1 / ENERGY BANDS IN SOLIDS

In this chapter we begin with a review of the basic atomic properties of matter leading to discrete electronic energy levels in atoms. We find that these energy levels are spread into energy bands in a crystal. This band structure allows us to distinguish between an insulator, a semiconductor, and a metal.

1-1 CHARGED PARTICLES

The charge, or quantity, of negative electricity and the mass of the electron have been found to be 1.60×10^{-19} C (coulomb) and 9.11×10^{-31} kg, respectively. The values of many important physical constants are given in Appendix A, and a list of conversion factors and prefixes is given in Appendix B. Some idea of the number of electrons per second that represents current of the usual order of magnitude is readily possible. For example, since the charge per electron is 1.60×10^{-19} C, the number of electrons per coulomb is the reciprocal of this number, or approximately, 6×10^{18} Further, since a current of 1 A (ampere) is the flow of 1 C/s, then a current of only 1 pA (1 picoampere, or 10^{-12} A) represents the motion of approximately 6 million electrons per second. Yet a current of 1 pA is so small that considerable difficulty is experienced in attempting to measure it.

The charge of a positive ion is an integral multiple of the charge of the electron, although it is of opposite sign. For the case of singly ionized particles, the charge is equal to that of the electron. For the case of doubly ionized particles, the ionic charge is twice that of the electron.

The mass of an atom is expressed as a number that is based on the choice of the atomic weight of oxygen equal to 16. The mass of a hypothetical atom of atomic weight unity is, by this definition, one-sixteenth that of the mass of monatomic oxygen and has been calculated to be 1.66×10^{-27} kg. Hence, *to calculate the mass in kilograms*

VISIT…

LANZAROTE

Caliente.COM

of any atom, it is necessary only to multiply the atomic weight of the atom by 1.66×10^{-27} *kg.* A table of atomic weights is given in Table 1-1 on p. 12.

The radius of the electron has been estimated as 10^{-15} m, and that of an atom as 10^{-10} m. These are so small that all charges are considered as mass points in the following sections.

In a semiconductor crystal such as silicon, two electrons are shared by each pair of ionic neighbors. Such a configuration is called a *covalent bond.* Under certain circumstances an electron may be missing from this structure, leaving a "hole" in the bond. These vacancies in the covalent bonds may move from ion to ion in the crystal and constitute a current equivalent to that resulting from the motion of free positive charges. The magnitude of the charge associated with the hole is that of a free electron. This very brief introduction to the concept of a hole as an effective charge carrier is elaborated upon in Chap. 2.

1-2 FIELD INTENSITY, POTENTIAL, ENERGY

By definition, *the force* $\mathbf{f}$ *(newtons) on a unit positive charge in an electric field is the electric field intensity* $\mathcal{E}$ *at that point.* Newton's second law determines the motion of a particle of charge q (coulombs), mass m (kilograms), moving with a velocity $\mathbf{v}$ (meters per second) in a field $\mathcal{E}$ (volts per meter).

$$\mathbf{f} = q\mathcal{E} = m\frac{d\mathbf{v}}{dt} \tag{1-1}$$

The mks (meter-kilogram-second) rationalized system of units is found to be most convenient for subsequent studies. Unless otherwise stated, this system of units is employed throughout this book.

Potential By definition, *the potential V (volts) of point B with respect to point A is the work done* against *the field* in taking a unit positive charge from A to B. This definition is valid for a three-dimensional field. For a one-dimensional problem with A at x_o and B at an arbitrary distance x, it follows that†

$$V \equiv -\int_{x_o}^{x} \mathcal{E}\,dx \tag{1-2}$$

where $\mathcal{E}$ now represents the X component of the field. Differentiating Eq. (1-2) gives

$$\mathcal{E} = -\frac{dV}{dx} \tag{1-3}$$

The minus sign shows that the electric field is directed from the region of higher potential to the region of lower potential. In three dimensions, the electric field equals the negative gradient of the potential.

---

† The symbol ≡ is used to designate "equal to by definition."

By definition, *the potential energy U (joules) equals the potential multiplied by the charge q under consideration*, or

$$U = qV \tag{1-4}$$

If an electron is being considered, q is replaced by $-q$ (where q is the *magnitude* of the electronic charge) and U has the same shape as V but is inverted.

The law of conservation of energy states that the total energy W, which equals the sum of the potential energy U and the kinetic energy $\frac{1}{2}mv^2$, remains constant. Thus, at any point in space,

$$W = U + \tfrac{1}{2}mv^2 = \text{constant} \tag{1-5}$$

As an illustration of this law, consider two parallel electrodes (A and B of Fig. 1-1*a*) separated a distance d, with B at a negative potential V_d with respect to A. An electron leaves the surface of A with a velocity v_o in the direction toward B. How much speed v will it have if it reaches B?

From the definition, Eq. (1-2), it is clear that only differences of potential have meaning, and hence let us arbitrarily ground A, that is, consider it to be at zero potential. Then the potential at B is $V = -V_d$, and the potential energy is $U = -qV = qV_d$. Equating the total energy at A to that at B gives

$$W = \tfrac{1}{2}mv_o^2 = \tfrac{1}{2}mv^2 + qV_d \tag{1-6}$$

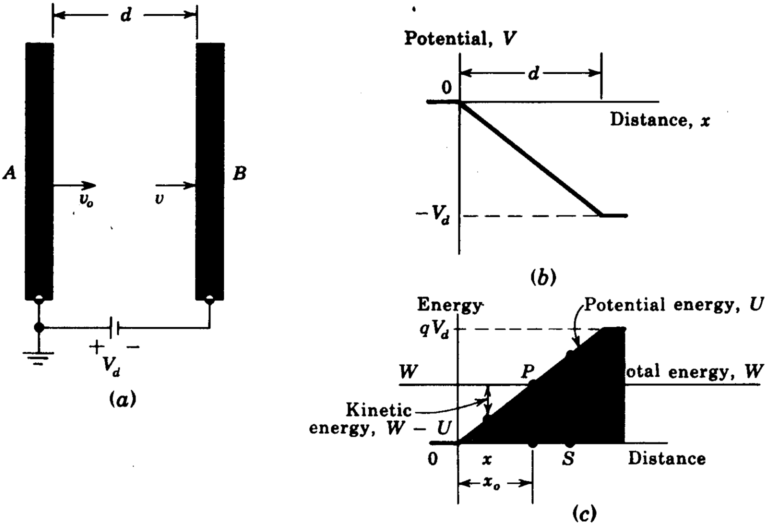


Fig. 1-1 (*a*) An electron leaves electrode A with an initial speed v_o and moves in a retarding field toward plate B; (*b*) the potential; (*c*) the potential-energy barrier between electrodes.

This equation indicates that v must be less than v_o, which is obviously correct since the electron is moving in a repelling field. Note that the final speed v attained by the electron in this conservative system is independent of the form of the variation of the field distribution between the plates and depends only upon the magnitude of the potential difference V_d. Also, if the electron is to reach electrode B, its initial speed must be large enough so that $\frac{1}{2}mv_o^2 > qV_d$. Otherwise, Eq. (1-6) leads to the impossible result that v is imaginary. We wish to elaborate on these considerations now.

The Concept of a Potential-energy Barrier For the configuration of Fig. 1-1*a* with electrodes which are large compared with the separation d, we can draw (Fig. 1-1*b*) a linear plot of potential V versus distance x (in the interelectrode space). The corresponding potential energy U versus x is indicated in Fig. 1-1*c*. Since potential is the potential energy per unit charge, curve *c* is obtained from curve *b* by multiplying each ordinate by the charge on the electron (a negative number). Since the total energy W of the electron remains constant, it is represented as a horizontal line. The kinetic energy at any distance x equals the difference between the total energy W and the potential energy U at this point. This difference is greatest at O, indicating that the kinetic energy is a maximum when the electron leaves the electrode A. At the point P this difference is zero, which means that no kinetic energy exists, so that the particle is at rest at this point. This distance x_o is the maximum that the electron can travel from A. At point P (where $x = x_o$) it comes momentarily to rest, and then reverses its motion and returns to A.

Consider a point such as S which is at a greater distance than x_o from electrode A. Here the total energy QS is less than the potential energy RS, so that the difference, which represents the kinetic energy, is negative. This is an impossible physical condition, however, since negative kinetic energy ($\frac{1}{2}mv^2 < 0$) implies an imaginary velocity. We must conclude that the particle can never advance a distance greater than x_o from electrode A.

The foregoing analysis leads to the very important conclusion that the shaded portion of Fig. 1-1*c* can never be penetrated by the electron. Thus, at point P, the particle acts *as if* it had collided with a solid wall, hill, or barrier and the direction of its flight had been altered. *Potential-energy barriers* of this sort play important role in the analyses of semiconductor devices.

It must be emphasized that the words "collides with" or "rebounds from" a potential "hill" are convenient descriptive phrases and that an actual encounter between two material bodies is not implied.

1-3 THE eV UNIT OF ENERGY

The joule (J) is the unit of energy in the mks system. In some engineering power problems this unit is very small, and a factor of 10^3 or 10^6 is introduced to convert from watts (1 W = 1 J/s) to kilowatts or megawatts, respectively.

However, in other problems, the joule is too large a unit, and a factor of 10^{-7} is introduced to convert from joules to ergs. For a discussion of the energies involved in electronic devices, even the erg is much too large a unit. This statement is not to be construed to mean that only minute amounts of energy can be obtained from electron devices. It is true that each electron possesses a tiny amount of energy, but as previously pointed out (Sec. 1-1), an enormous number of electrons are involved even in a small current, so that considerable power may be represented.

A unit of work or energy, called the *electron volt* (eV), is defined as follows:

$$1 \text{ eV} \equiv 1.60 \times 10^{-19} \text{ J}$$

Of course, any type of energy, whether it be electric, mechanical, thermal, etc., may be expressed in electron volts.

The name *electron volt* arises from the fact that, if an electron falls through a potential of one volt, its kinetic energy will increase by the decrease in potential energy, or by

$$qV = (1.60 \times 10^{-19} \text{ C})(1 \text{ V}) = 1.60 \times 10^{-19} \text{ J} = 1 \text{ eV}$$

However, as mentioned above, the electron-volt unit may be used for any type of energy, and is not restricted to problems involving electrons.

A potential-energy barrier of E (electron volts) is equivalent to a potential hill of V (volts) if these quantities are related by

$$qV = 1.60 \times 10^{-19} E \tag{1-7}$$

Note that V and E are *numerically* identical but dimensionally different.

1-4 THE NATURE OF THE ATOM

We wish to develop the band structure of a solid, which will allow us to distinguish between an insulator, a semiconductor, and a metal. We begin with a review of the basic properties of matter leading to discrete electronic energy levels in atoms.

Rutherford, in 1911, found that the atom consists of a nucleus of positive charge that contains nearly all the mass of the atom. Surrounding this central positive core are negatively charged electrons. As a specific illustration of this atomic model, consider the hydrogen atom. This atom consists of a positively charged nucleus (a proton) and a single electron. The charge on the proton is positive and is equal in magnitude to the charge on the electron. Therefore the atom as a whole is electrically neutral. Because the proton carries practically all the mass of the atom, it will remain substantially immobile, whereas the electron will move about it in a closed orbit. The force of attraction between the electron and the proton follows Coulomb's law. It can be shown from classical mechanics that the resultant closed path will be a circle or an ellipse under the action of such a force. This motion is exactly analogous to that of the planets about the sun, because in both cases the force varies inversely as the square of the distance between the particles.

Assume, therefore, that the orbit of the electron in this planetary model of the atom is a circle, the nucleus being supposed fixed in space. It is a simple matter to calculate its radius in terms of the total energy W of the electron. The force of attraction between the nucleus and the electron of the hydrogen atom is $q^2/4\pi\epsilon_o r^2$, where the electronic charge q is in coulombs, the separation r between the two particles is in meters, the force is in newtons, and ϵ_o is the permittivity of free space.† By Newton's second law of motion, this must be set equal to the product of the electronic mass m in kilograms and the acceleration v^2/r toward the nucleus, where v is the speed of the electron in its circular path, in meters per second. Then

$$\frac{q^2}{4\pi\epsilon_o r^2} = \frac{mv^2}{r} \tag{1-8}$$

Furthermore, the potential energy of the electron at a distance r from the nucleus is $-q^2/4\pi\epsilon_o r$, and its kinetic energy is $\frac{1}{2}mv^2$. Then, according to the conservation of energy,

$$W = \tfrac{1}{2}mv^2 - \frac{q^2}{4\pi\epsilon_o r} \tag{1-9}$$

where the energy is in joules. Combining this expression with (1-8) produces

$$W = -\frac{q^2}{8\pi\epsilon_o r} \tag{1-10}$$

which gives the desired relationship between the radius and the energy of the electron. This equation shows that the total energy of the electron is always negative. The negative sign arises because the potential energy has been chosen to be zero when r is infinite. This expression also shows that the energy of the electron becomes smaller (i.e., more negative) as it approaches closer to the nucleus.

The foregoing discussion of the planetary atom has been considered only from the point of view of classical mechanics. However, an accelerated charge must radiate energy, in accordance with the classical laws of electromagnetism. If the charge is performing oscillations of a frequency f, the radiated energy will also be of this frequency. Hence, classically, it must be concluded that the frequency of the emitted radiation equals the frequency with which the electron is rotating in its circular orbit.

There is one feature of this picture that cannot be reconciled with experiment. If the electron is radiating energy, its total energy must decrease by the amount of this emitted energy. As a result the radius r of the orbit must decrease, in accordance with Eq. (1-10). Consequently, as the atom radiates energy, the electron must move in smaller and smaller orbits, eventually falling into the nucleus. Since the frequency of oscillation depends upon the size of the circular orbit, the energy radiated would be of a gradually changing frequency. Such a conclusion, however, is incompatible with the sharply defined frequencies of spectral lines.

---

† The numerical value of ϵ_o is given in Appendix A.

The Bohr Atom The difficulty mentioned above was resolved by Bohr in 1913. He postulated the following three fundamental laws:

1. Not all energies as given by classical mechanics are possible, but the atom can possess only certain discrete energies. While in states corresponding to these discrete energies, the electron does *not* emit radiation, and the electron is said to be in a *stationary*, or nonradiating, state.

2. In a transition from one stationary state corresponding to a definite energy W_2 to another stationary state, with an associated energy W_1, radiation will be emitted. The frequency of this radiant energy is given by

$$f = \frac{W_2 - W_1}{h} \tag{1-11}$$

where h is Planck's constant in joule-seconds, the W's are expressed in joules, and f is in cycles per second, or hertz.

3. A stationary state is determined by the condition that the angular momentum of the electron in this state is quantized and must be an integral multiple of $h/2\pi$. Thus

$$mvr = \frac{nh}{2\pi} \tag{1-12}$$

where n is an integer.

Combining Eqs. (1-8) and (1-12), we obtain the radii of the stationary states (Prob. 1-13), and from Eq. (1-10) the energy level in joules of each state is found to be

$$W_n = -\frac{mq^4}{8h^2\epsilon_o{}^2}\frac{1}{n^2} \tag{1-13}$$

Then, upon making use of Eq. (1-11), the exact frequencies found in the hydrogen spectrum are obtained—a remarkable achievement. The radius of the lowest state is found to be 0.5 Å.

1-5 ATOMIC ENERGY LEVELS

For each integral value of n in Eq. (1-13) a horizontal line is drawn. These lines are arranged vertically in accordance with the numerical values calculated from Eq. (1-13). Such a convenient pictorial representation is called an *energy-level diagram* and is indicated in Fig. 1-2 for hydrogen. The number to the left of each line gives the energy of this level in electron volts. The number immediately to the right of a line is the value of n. Theoretically, an infinite number of levels exist for each atom, but only the first five and the

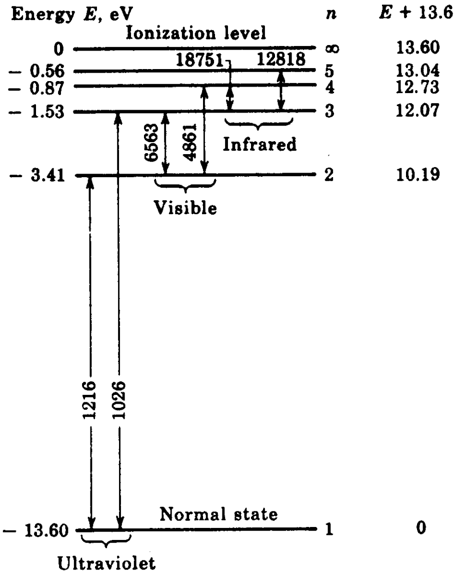


Fig. 1-2 The lowest five energy levels and the ionization level of hydrogen. The spectral lines are in angstrom units.

level for $n = \infty$ are indicated in Fig. 1-2. The horizontal axis has no significance here, but in extending such energy-level diagrams to solids, the X axis will be used to represent the separation of atoms within a crystal (Fig. 1-3) or the distance within a solid. In such cases the energy levels are not constant, but rather are functions of x.

It is customary to express the energy value of the stationary states in electron volts E rather than in joules W. Also, it is more common to specify the emitted radiation by its wavelength λ in angstroms rather than by its frequency f in hertz. In these units, Eq. (1-11) may be rewritten in the form

$$\lambda = \frac{12{,}400}{E_2 - E_1} \tag{1-14}$$

Since only differences of energy enter into this expression, the zero state may be chosen at will. It is convenient and customary to choose the lowest energy state as the zero level. Such a normalized scale is indicated to the extreme right in Fig. 1-2. The lowest energy state is called the *normal*, or *ground*, level, and the other stationary states of the atom are called *excited*, *radiating*, *critical*, or *resonance*, levels.

As the electron is given more and more energy, it moves into stationary states which are farther and farther away from the nucleus. When its energy is large enough to move it completely out of the field of influence of the ion, it becomes "detached" from it. The energy required for this process to occur is called the *ionization potential* and is represented as the highest state in the energy-level diagram; 13.60 eV for hydrogen.

Collisions of Electrons with Atoms The foregoing discussion shows that energy must be supplied to an atom in order to excite or ionize the atom. One of the most important ways to supply this energy is by electron impact. Suppose that an electron is accelerated by the potential applied to a discharge tube. The energy gained from the field may then be transferred to an atom when the electron collides with the atom. If the bombarding electron has gained more than the requisite energy from the discharge to raise the atom from its normal state to a particular resonance level, the amount of energy in excess of that required for excitation will be retained by the incident electron as kinetic energy after the collision.

If an impinging electron possesses an amount of energy at least equal to the ionization potential of the gas, it may deliver this energy to an electron of the atom and completely remove it from the parent atom. Three charged particles result from such an ionizing collision: two electrons and a positive ion.

The Photon Nature of Light Assume that an atom has been raised from the ground state to an excited level by electron bombardment. The mean life of an excited state ranges from 10^{-7} to 10^{-10} s, the excited electron returning to its previous state after the lapse of this time. In this transition, the atom must lose an amount of energy equal to the difference in energy between the two states that it has successively occupied, this energy appearing in the form of radiation. According to the postulates of Bohr, this energy is emitted in the form of a photon of light, the frequency of this radiation being given by Eq. (1-11), or the wavelength by Eq. (1-14). The term *photon* denotes an amount of radiant energy equal to the constant h times the frequency. This quantized nature of an electromagnetic wave was first introduced by Planck, in 1901, in order to verify theoretically the blackbody radiation formula obtained experimentally.

The photon concept of radiation may be difficult to comprehend at first. Classically, it was believed that the atoms were systems that emitted radiation *continuously* in all directions. According to the foregoing theory, however, this is not true, the emission of light by an atom being a discontinuous process. That is, the atom radiates only when it makes a transition from one energy level to a lower energy state. In this transition, it emits a definite amount of energy of one particular frequency, namely, one photon hf of light. Of course, when a luminous discharge is observed, this discontinuous nature of radiation is not suspected because of the enormous number of atoms that are radiating energy and, correspondingly, because of the immense number of photons that are emitted in unit time.

Spectral Lines The arrows in Fig. 1-2 represent six possible transitions between stationary states. The attached number gives the wavelength of the emitted radiation. For example, the ultraviolet line 1,216 Å is radiated when the hydrogen atom drops from its first excited state, $n = 2$, to its normal state, $n = 1$.

Another important method, called *photoexcitation*, by which an atom may be elevated into an excited energy state, is to have radiation fall on the gas. An atom may absorb a photon of frequency f and thereby move from the level of energy W_1 to the high energy level W_2, where $W_2 = W_1 + hf$. An extremely important feature of excitation by photon capture is that *the photon will not be absorbed unless its energy corresponds exactly to the energy difference between two stationary levels of the atom with which it collides.* For example, if a normal hydrogen atom is to be raised to its first excited state by means of radiation, the wavelength of this light must be 1,216 Å (which is in the ultraviolet region of the spectrum).

When a photon is absorbed by an atom, the excited atom may return to its normal state in one jump, or it may do so in several steps. If the atom falls into one or more excitation levels before finally reaching the normal state, it will emit several photons. These will correspond to energy differences between the successive excited levels into which the atom falls. None of the emitted photons will have the frequency of the absorbed radiation! This *fluorescence* cannot be explained by classical theory, but is readily understood once Bohr's postulates are accepted.

Photoionization If the frequency of the impinging photon is sufficiently high, it may have enough energy to ionize the atom. The photon vanishes with the appearance of an electron and a positive ion. Unlike the case of photoexcitation, the photon need not possess an energy corresponding exactly to the ionization energy of the atom. It need merely possess *at least* this much energy. If it possesses more than ionizing energy, the excess will appear as the kinetic energy of the emitted electron and positive ion. It is found by experiment, however, that the maximum probability of photoionization occurs when the energy of the photon is equal to the ionization potential, the probability decreasing rapidly for higher photon energies.

Wave Mechanics Since a photon is absorbed by only one atom, the photon acts as if it were concentrated in a very small volume of space, in contradiction to the concept of a wave associated with radiation. De Broglie, in 1924, postulated that the dual character of wave and particle is not limited to radiation, but is also exhibited by particles such as electrons, atoms, or macroscopic masses. He postulated that a particle of momentum $p = mv$ has a wavelength λ associated with it given by

$$\lambda = \frac{h}{p} \tag{1-15}$$

We can make use of the wave properties of a moving electron to establish Bohr's postulate that a stationary state is determined by the condition that the angular momentum must be an integral multiple of $h/2\pi$. It seems reasonable to assume that an orbit of radius r will correspond to a stationary state if it contains a standing-wave pattern. In other words, a stable orbit is one whose circumference is exactly equal to the electronic wavelength λ, or to $n\lambda$, where n is an integer (but not zero). Thus

$$2\pi r = n\lambda = \frac{nh}{mv} \tag{1-16}$$

Clearly, Eq. (1-16) is identical with the Bohr condition [Eq. (1-12)].

Schrödinger carried the implication of the wave nature of matter further and developed a wave equation to describe electron behavior in a potential field $U(x, y, z)$. The solution of this differential equation is called the wave function, and it determines the probability density at each point in space of finding the electron with total energy W. If the potential energy, $U = -q^2/4\pi\epsilon_0 r$, for the electron in the hydrogen atom is substituted into the Schrödinger equation, it is found that a meaningful physical solution is possible only if W is given by precisely the energy levels in Eq. (1-13), which were obtained from the simpler Bohr picture of the atom.

1-6 ELECTRONIC STRUCTURE OF THE ELEMENTS

The solution of the Schrödinger equation for hydrogen or any multielectron atom requires three quantum numbers. These are designated by n, l, and m_l and are restricted to the following integral values:

$$\begin{aligned} n &= 1, 2, 3, \ldots \\ l &= 0, 1, 2, \ldots, (n-1) \\ m_l &= 0, \pm 1, \pm 2, \ldots, \pm l \end{aligned}$$

To specify a wave function completely it is found necessary to introduce a fourth quantum number. This *spin* quantum number m_s may assume only two values, $+\frac{1}{2}$ or $-\frac{1}{2}$ (corresponding to the same energy).

The Exclusion Principle The periodic table of the chemical elements (given in Table 1-1) may be explained by invoking a law enunciated by Pauli in 1925. He stated that *no two electrons in an electronic system can have the same set of four quantum numbers, n, l, m_l, and m_s.* This statement that no two electrons may occupy the same quantum state is known as the *Pauli exclusion principle.*

Electronic Shells All the electrons in an atom which have the same value of n are said to belong to the same *electron shell.* These shells are identified by the letters K, L, M, N, . . . , corresponding to $n = 1, 2, 3, 4, \ldots$, respectively. A shell is divided into *subshells* corresponding to different values of l and identified as s, p, d, f, . . . , corresponding to $l = 0, 1, 2, 3, \ldots$, respectively. Taking account of the exclusion principle, the distribution of

TABLE 1-1 Periodic table of the elements†

Period	Group IA	Group IIA	Group IIIB	Group IVB	Group VB	Group VIB	Group VIIB	Group VIII			Group IB	Group IIB	Group IIIA	Group IVA	Group VA	Group VIA	Group VIIA	Inert gases
1	H 1 1.01																	He 2 4.00
2	Li 3 6.94	Be 4 9.01											B 5 10.81	C 6 12.01	N 7 14.01	O 8 16.00	F 9 19.00	Ne 10 20.18
3	Na 11 22.99	Mg 12 24.31											Al 13 26.98	Si 14 28.09	P 15 30.97	S 16 32.06	Cl 17 35.45	Ar 18 39.95
4	K 19 39.10	Ca 20 40.08	Sc 21 44.96	Ti 22 47.90	V 23 50.94	Cr 24 52.00	Mn 25 54.94	Fe 26 55.85	Co 27 58.93	Ni 28 58.71	Cu 29 63.54	Zn 30 65.37	Ga 31 69.72	Ge 32 72.59	As 33 74.92	Se 34 78.96	Br 35 79.91	Kr 36 83.80
5	Rb 37 85.47	Sr 38 87.62	Y 39 88.90	Zr 40 91.22	Nb 41 92.91	Mo 42 95.94	Tc 43 (99)	Ru 44 101.07	Rh 45 102.90	Pd 46 106.4	Ag 47 107.87	Cd 48 112.40	In 49 114.82	Sn 50 118.69	Sb 51 121.75	Te 52 127.60	I 53 126.90	Xe 54 131.30
6	Cs 55 132.90	Ba 56 137.34	La 57 138.91	Hf 72 178.49	Ta 73 180.95	W 74 183.85	Re 75 186.2	Os 76 190.2	Ir 77 192.2	Pt 78 195.09	Au 79 196.97	Hg 80 200.59	Tl 81 204.37	Pb 82 207.19	Bi 83 208.98	Po 84 (210)	At 85 (210)	Rn 86 (222)
7	Fr 87 (223)	Ra 88 (226)	Ac 89 (227)	Th 90 232.04	Pa 91 (231)	U 92 238.04	Np 93 (237)	Pu 94 (242)	Am 95 (243)	Cm 96 (247)	Bk 97 (247)	Cf 98 (251)	Es 99 (254)	Fm 100 (253)	Nd 101 (256)	No 102 (254)	Lw 103 (257)	

The Rare Earths

Ce 58 140.12	Pr 59 140.91	Nd 60 144.24	Pm 61 (147)	Sm 62 150.35	Eu 63 151.96	Gd 64 157.25	Tb 65 158.92	Dy 66 162.50	Ho 67 164.93	Er 68 167.26	Tm 69 168.93	Yb 70 173.04	Lu 71 174.97

† The number to the right of the symbol for the element gives the atomic number. The number below the symbol for the element gives the atomic weight.

TABLE 1-2 Electron shells and subshells

Shell	K	L		M			N			
n	1	2		3			4			
l	0	0	1	0	1	2	0	1	2	3
Subshell	s	s	p	s	p	d	s	p	d	f
m_l	0	0	$0, \pm 1$	0	$0, \pm 1$	$0, \pm 1, \pm 2$	0	$0, \pm 1$	$0, \pm 1, \pm 2$	$0, \ldots, \pm 3$
Number of electrons	2	2	6	2	6	10	2	6	10	14
	2	8		18			32			

electrons in an atom among the shells and subshells is indicated in Table 1-2. Actually, seven shells are required to account for all the chemical elements, but only the first four are indicated in the table.

There are two states for $n = 1$ corresponding to $l = 0$, $m_l = 0$, and $m_s = \pm\frac{1}{2}$. These are called the 1s states. There are two states corresponding to $n = 2$, $l = 0$, $m_l = 0$, and $m_s = \pm\frac{1}{2}$. These constitute the 2s subshell. There are, in addition, six energy levels corresponding to $n = 2$, $l = 1$, $m_l = -1, 0,$ or $+1$, and $m_s = \pm\frac{1}{2}$. These are designated as the 2p subshell. Hence, as indicated in Table 1-2, the total number of electrons in the L shell is $2 + 6 = 8$. In a similar manner we may verify that a d subshell contains a maximum of 10 electrons, an f subshell a maximum of 14 electrons, etc.

The atomic number Z gives the number of electrons orbiting about the nucleus. Let us use superscripts to designate the number of electrons in a particular subshell. Then sodium, Na, for which $Z = 11$, has an electronic configuration designated by $1s^2 2s^2 2p^6 3s^1$. Note that Na has a single electron in the outermost unfilled subshell, and hence is said to be monovalent. This same property is possessed by all the alkali metals (Li, Na, K, Rb, and Cs), which accounts for the fact that these elements in the same group in the periodic table (Table 1-1) have similar chemical properties.

The inner-shell electrons are very strongly bound to an atom, and cannot be easily removed. That is, the electrons closest to the nucleus are the most tightly bound, and so have the lowest energy. Also, atoms for which the electrons exist in closed shells form very stable configurations. For example, the inert gases He, Ne, A, Kr, and Xe, all have either completely filled shells or, at least, completely filled subshells.

Carbon, silicon, germanium, and tin have the electronic configurations indicated in Table 1-3. Note that each of these elements has completely filled subshells except for the outermost p shell, which contains only two of the six possible electrons. Despite this similarity, carbon in crystalline form (diamond) is an insulator, silicon and germanium solids are semiconductors, and tin is a metal. This apparent anomaly is explained in the next section.

TABLE 1-3 Electronic configuration in Group IVA

Element	Atomic number	Configuration
C	6	$1s^22s^22p^2$
Si	14	$1s^22s^22p^63s^23p^2$
Ge	32	$1s^22s^22p^63s^23p^63d^{10}4s^24p^2$
Sn	50	$1s^22s^22p^63s^23p^63d^{10}4s^24p^64d^{10}5s^25p^2$

1-7 THE ENERGY-BAND THEORY OF CRYSTALS

X-ray and other studies reveal that most metals and semiconductors are crystalline in structure. A crystal consists of a space array of atoms or molecules (strictly speaking, ions) built up by regular repetition in three dimensions of some fundamental structural unit. The electronic energy levels discussed for a single free atom (as in a gas, where the atoms are sufficiently far apart not to exert any influence on one another) do not apply to the same atom in a crystal. This is so because the potential characterizing the crystalline structure is now a periodic function in space whose value at any point is the result of contributions from every atom. When atoms form crystals, it is found that the energy levels of the inner-shell electrons are not affected appreciably by the presence of the neighboring atoms. However, the levels of the outer-shell electrons are changed considerably, since these electrons are shared by more than one atom in the crystal. The new energy levels of the outer electrons can be determined by means of quantum mechanics, and it is found that coupling between the outer-shell electrons of the atoms results in a *band* of closely spaced energy states, instead of the widely separated energy levels of the isolated atom (Fig. 1-3). A qualitative discussion of this energy-band structure follows.

Consider a crystal consisting of N atoms of one of the elements in Table 1-3. Imagine that it is possible to vary the spacing between atoms without altering the type of fundamental crystal structure. If the atoms are so far apart that the interaction between them is negligible, the energy levels will coincide with those of the isolated atom. The outer two subshells for each element in Table 1-3 contain two s electrons and two p electrons. Hence, if we ignore the inner-shell levels, then, as indicated to the extreme right in Fig. 1-3a, there are $2N$ electrons completely filling the $2N$ possible s levels, all at the same energy. Since the p atomic subshell has six possible states, our imaginary crystal of widely spaced atoms has $2N$ electrons, which fill only one-third of the $6N$ possible p states, all at the same level.

If we now decrease the interatomic spacing of our imaginary crystal (moving from right to left in Fig. 1-3a), an atom will exert an electric force on its neighbors. Because of this coupling between atoms, the atomic-wave

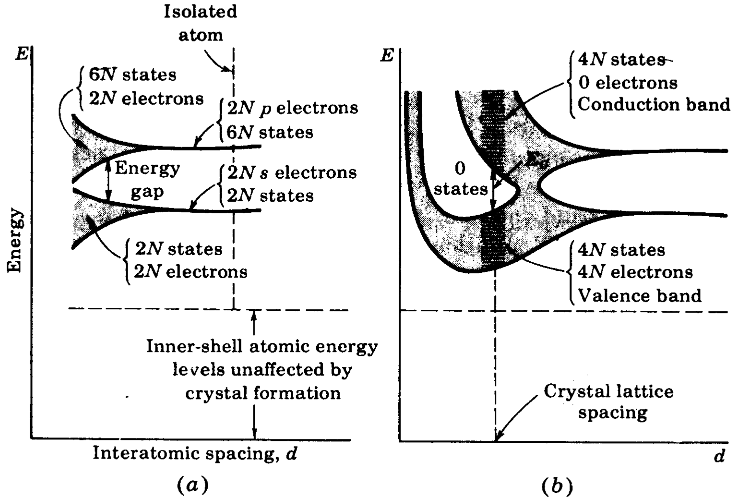

Isolated atom
E
Energy
6N states
2N electrons
2N p electrons
6N states
Energy gap
2N s electrons
2N states
2N states
2N electrons
Inner-shell atomic energy levels unaffected by crystal formation
Interatomic spacing, d
(a)
E
4N states
0 electrons
Conduction band
0 states
E_g
4N states
4N electrons
Valence band
Crystal lattice spacing
d
(b)

Fig. 1-3 Illustrating how the energy levels of isolated atoms are split into energy bands when these atoms are brought into close proximity to form a crystal.

functions overlap, and the crystal becomes an electronic *system* which must obey the Pauli exclusion principle. Hence the $2N$ degenerate s states must spread out in energy. The separation between levels is small, but since N is very large ($\sim 10^{23}$ cm^{-3}), the total spread between the minimum and maximum energy may be several electron volts if the interatomic distance is decreased sufficiently. This large number of discrete but closely spaced energy levels is called an *energy band*, and is indicated schematically by the lower shaded region in Fig. 1-3*a*. The $2N$ states in this band are completely filled with $2N$ electrons. Similarly, the upper shaded region in Fig. 1-3*a* is a band of $6N$ states which has only $2N$ of its levels occupied by electrons.

Note that there is an energy gap (a forbidden band) between the two bands discussed above and that this gap decreases as the atomic spacing decreases. For small enough distances (not indicated in Fig. 1-3*a* but shown in Fig. 1-3*b*) these bands will overlap. Under such circumstances the $6N$ upper states merge with the $2N$ lower states, giving a total of $8N$ levels, half of which are occupied by the $2N + 2N = 4N$ available electrons. At this spacing each atom has given up four electrons to the band; these electrons can no longer be said to orbit in s or p subshells of an isolated atom, but rather they belong to the crystal as a whole. In this sense the elements in Table 1-3 are tetravalent, since they contribute four electrons each to the crystal. The band these electrons occupy is called the *valence band.*

If the spacing between atoms is decreased below the distance at which the bands overlap, the interaction between atoms is indeed large. The energy-band structure then depends upon the orientation of the atoms relative to one another in space (the crystal structure) and upon the atomic number, which determines the electrical constitution of each atom. Solutions of Schrödinger's equation are complicated and have been obtained approximately for only relatively few crystals. These solutions lead us to expect an energy-band diagram somewhat as pictured[1] in Fig. 1-3*b*. At the crystal-lattice spacing (the dashed vertical line), we find the valence band *filled* with $4N$ electrons separated by a forbidden band (no allowed energy states) of extent E_G from an *empty* band consisting of $4N$ additional states. This upper vacant band is called the *conduction band*, for reasons given in the next section.

1-8 INSULATORS, SEMICONDUCTORS, AND METALS

A very poor conductor of electricity is called an *insulator;* an excellent conductor is a *metal;* and a substance whose conductivity lies between these extremes is a *semiconductor.* A material may be placed in one of these three classes, depending upon its energy-band structure.

Insulator The energy-band structure of Fig. 1-3*b* at the normal lattice spacing is indicated schematically in Fig. 1-4*a*. For a diamond (carbon) crystal the region containing no quantum states is several electron volts high ($E_G \approx 6$ eV). This large forbidden band separates the filled valence region from the vacant conduction band. The energy which can be supplied to an electron from an applied field is too small to carry the particle from the filled into the vacant band. Since the electron cannot acquire sufficient applied energy, conduction is impossible, and hence diamond is an *insulator.*

Semiconductor A substance for which the width of the forbidden energy region is relatively small (~1 eV) is called a *semiconductor.* Graphite, a

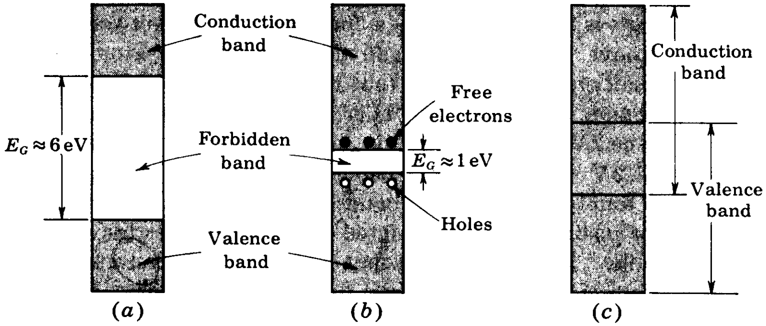


Fig. 1-4 Energy-band structure of (*a*) an insulator, (*b*) a semiconductor, and (*c*) a metal.

crystalline form of carbon but having a crystal symmetry which is different from diamond, has such a small value of E_G, and it is a semiconductor. The most important practical semiconductor materials are germanium and silicon, which have values of E_G of 0.785 and 1.21 eV, respectively, at 0°K. Energies of this magnitude normally cannot be acquired from an applied field. Hence the valence band remains full, the conduction band empty, and these materials are insulators at low temperatures. However, the conductivity increases with temperature, as we explain below. These substances are known as *intrinsic (pure) semiconductors.*

As the temperature is increased, some of these valence electrons acquire *thermal* energy greater than E_G, and hence move into the conduction band. These are now free electrons in the sense that they can move about under the influence of even a small applied field. These free, or conduction, electrons are indicated schematically by dots in Fig. 1-4*b*. The insulator has now become slightly conducting; it is a *semiconductor.* The absence of an electron in the valence band is represented by a small circle in Fig. 1-4*b*, and is called a *hole.* The phrase "holes in a semiconductor" therefore refers to the empty energy levels in an otherwise filled valence band.

The importance of the hole is that it may serve as a carrier of electricity, comparable in effectiveness with the free electron. The mechanism by which a hole contributes to conductivity is explained in Sec. 2-2. We also show in Chap. 2 that if certain impurity atoms are introduced into the crystal, these result in allowable energy states which lie in the forbidden energy gap. We find that these impurity levels also contribute to the conduction. A semiconductor material where this conduction mechanism predominates is called an *extrinsic (impurity) semiconductor.*

Since the band-gap energy of a crystal is a function of interatomic spacing (Fig. 1-3), it is not surprising that E_G depends somewhat on temperature. It has been determined experimentally that E_G decreases with temperature, and this dependence is given in Sec. 2-5.

Metal A solid which contains a partly filled band structure is called a *metal.* Under the influence of an applied electric field the electrons may acquire additional energy and move into higher states. Since these mobile electrons constitute a current, this substance is a conductor and the partly filled region is the conduction band. One example of the band structure of a metal is given in Fig. 1-4*c*, which shows overlapping valence and conduction bands.

REFERENCES

1. Adler, R. B., A. C. Smith, and R. L. Longini: "Introduction to Semiconductor Physics," vol. 1, p. 78, Semiconductor Electronics Education Committee, John Wiley & Sons, Inc., New York, 1964.

2. Shockley, W.: "Electrons and Holes in Semiconductors," D. Van Nostrand Company, Inc., Princeton, N.J., 1963.

REVIEW QUESTIONS

1-1 Define *potential energy* in words and as an equation.

1-2 Define an *electron volt.*

1-3 State Bohr's three postulates for the atom.

1-4 Define a *photon.*

1-5 Define (*a*) *photoexcitation;* (*b*) *photoionization.*

1-6 State the *Pauli exclusion principle.*

1-7 Give the electronic configuration for an atom of a specified atomic number Z. For example, $Z = 32$ for germanium.

1-8 Explain why the energy levels of an atom become energy bands in a solid.

1-9 What is the difference between the band structure of an insulator and of a semiconductor?

1-10 What is the difference between the band structure of a semiconductor and of a metal?

1-11 Explain why a semiconductor acts as an insulator at 0°K and why its conductivity increases with increasing temperature.

1-12 What is the distinction between an intrinsic and an extrinsic semiconductor?

1-13 Define a *hole* in a semiconductor.

2 / TRANSPORT PHENOMENA IN SEMICONDUCTORS

The current in a metal is due to the flow of negative charges (*electrons*), whereas the current in a semiconductor results from the movement of both electrons and positive charges (*holes*). A semiconductor may be doped with impurity atoms so that the current is due predominantly either to electrons or to holes. The transport of the charges in a crystal under the influence of an electric field (a *drift* current), and also as a result of a nonuniform concentration gradient (a *diffusion* current), is investigated.

2-1 MOBILITY AND CONDUCTIVITY

In the preceding chapter we presented an energy-band picture of metals, semiconductors, and insulators. In a metal the outer, or valence, electrons of an atom are as much associated with one ion as with another, so that the electron attachment to any individual atom is almost zero. In terms of our previous discussion this means that the band occupied by the valence electrons may not be completely filled and that there are no forbidden levels at higher energies. Depending upon the metal, at least one, and sometimes two or three, electrons per atom are free to move throughout the interior of the metal under the action of applied fields.

Figure 2-1 is a two-dimensional schematic picture of the charge distribution within a metal. The shaded regions represent the net positive charge of the nucleus and the tightly bound inner electrons. The black dots represent the outer, or valence, electrons in the atom. It is these electrons that cannot be said to belong to any particular atom; instead, they have completely lost their individuality and can wander freely about from atom to atom in the metal. Thus a metal is

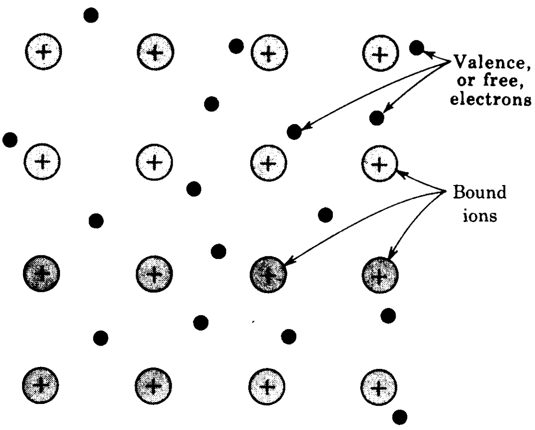


Fig. 2-1 Schematic arrangement of the atoms in one plane in a metal, drawn for monovalent atoms. The black dots represent the electron gas, each atom having contributed one electron to this gas.

visualized as a region containing a periodic three-dimensional array of heavy, tightly bound ions permeated with a swarm of electrons that may move about quite freely. This picture is known as the *electron-gas* description of a metal.

According to the electron-gas theory of a metal, the electrons are in continuous motion, the direction of flight being changed at each collision with the heavy (almost stationary) ions. The average distance between collisions is called the *mean free path.* Since the motion is random, then, on an average, there will be as many electrons passing through unit area in the metal in any direction as in the opposite direction in a given time. Hence the average current is zero.

Let us now see how the situation is changed if a constant electric field $\mathcal{E}$ (volts per meter) is applied to the metal. As a result of this electrostatic force, the electrons would be accelerated and the velocity would increase indefinitely with time, were it not for the collisions with the ions. However, at each inelastic collision with an ion, an electron loses energy, and a steady-state condition is reached where a finite value of *drift speed* v is attained.[1] This drift velocity is in the direction opposite to that of the electric field. The speed at a time t between collision is at, where $a = q\mathcal{E}/m$ is the acceleration. Hence the average speed v is proportional to $\mathcal{E}$. Thus

$$v = \mu\mathcal{E} \tag{2-1}$$

where μ (square meters per volt-second) is called the *mobility* of the electrons.

According to the foregoing theory, a steady-state drift speed has been superimposed upon the random thermal motion of the electrons. Such a directed flow of electrons constitutes a current. We now calculate the magnitude of the current.

Current Density If N electrons are contained in a length L of conductor (Fig. 2-2), and if it takes an electron a time T s to travel a distance of L m in

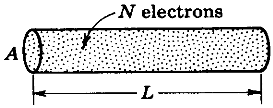


Fig. 2-2 Pertaining to the calculation of current density.

the conductor, the total number of electrons passing through any cross section of wire in unit time is N/T. Thus the total charge per second passing any area, which, by definition, is the current in amperes, is

$$I = \frac{Nq}{T} = \frac{Nqv}{L} \tag{2-2}$$

because L/T is the average, or *drift*, speed v m/s of the electrons. By definition, the current density, denoted by the symbol J, is the current per unit area of the conducting medium. That is, assuming a uniform current distribution,

$$J \equiv \frac{I}{A} \tag{2-3}$$

where J is in amperes per square meter, and A is the cross-sectional area (in meters) of the conductor. This becomes, by Eq. (2-2),

$$J = \frac{Nqv}{LA} \tag{2-4}$$

From Fig. 2-2 it is evident that LA is simply the volume containing the N electrons, and so N/LA is the electron concentration n (in electrons per cubic meter). Thus

$$n = \frac{N}{LA} \tag{2-5}$$

and Eq. (2-4) reduces to

$$J = nqv = \rho v \tag{2-6}$$

where $\rho \equiv nq$ is the charge density, in coulombs per cubic meter, and v is in meters per second.

This derivation is independent of the form of the conducting medium. Consequently, Fig. 2-2 does not necessarily represent a wire conductor. It may represent equally well a portion of a gaseous-discharge tube or a volume element of a semiconductor. Furthermore, neither ρ nor v need be constant, but may vary from point to point in space or may vary with time.

Conductivity From Eqs. (2-6) and (2-1)

$$J = nqv = nq\mu\mathcal{E} = \sigma\mathcal{E} \tag{2-7}$$

where

$$\sigma = nq\mu \tag{2-8}$$

is the *conductivity* of the metal in (ohm-meter)$^{-1}$. Equation (2-7) is recognized as Ohm's law, namely, the conduction current is proportional to the applied voltage. As already mentioned, the energy which the electrons acquire from the applied field is, as a result of collisions, given to the lattice ions. Hence power is dissipated within the metal by the electrons, and the power density (Joule heat) is given by $J\mathcal{E} = \sigma\mathcal{E}^2$ (watts per cubic meter).

2-2 ELECTRONS AND HOLES IN AN INTRINSIC SEMICONDUCTOR[1]

From Eq. (2-8) we see that the conductivity is proportional to the concentration n of free electrons. For a good conductor, n is very large ($\sim 10^{28}$ electrons/m^3); for an insulator, n is very small ($\sim 10^7$); and for a semiconductor, n lies between these two values. The valence electrons in a semiconductor are not free to wander about as they are in a metal, but rather are trapped in a bond between two adjacent ions, as explained below.

The Covalent Bond Germanium and silicon are the two most important semiconductors used in electronic devices. The crystal structure of these materials consists of a regular repetition in three dimensions of a unit cell having the form of a tetrahedron with an atom at each vertex. This structure is illustrated symbolically in two dimensions in Fig. 2-3. Germanium has a total of 32 electrons in its atomic structure, arranged in shells as indicated in Table 1-3. As explained in Sec. 1-7, each atom in a germanium crystal contributes four valence electrons, so that the atom is tetravalent. The inert ionic core of the germanium atom carries a positive charge of +4 measured in units of the electronic charge. The binding forces between neighboring atoms result from the fact that each of the valence electrons of a germanium atom is shared by one of its four nearest neighbors. This *electron-pair*, or *covalent, bond* is represented in Fig. 2-3 by the two dashed lines which join each atom to each of its neighbors. The fact that the valence electrons serve to bind one atom to the next also results in the valence electron being tightly bound to the

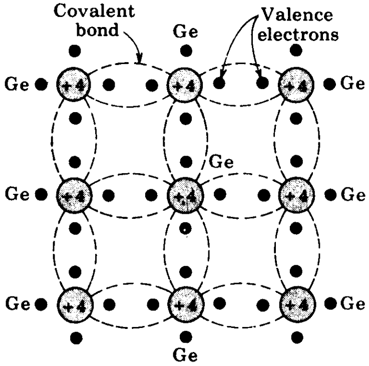


Fig. 2-3 Crystal structure of germanium, illustrated symbolically in two dimensions.

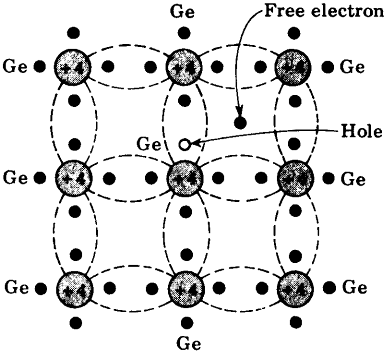


Fig. 2-4 Germanium crystal with a broken covalent bond.

nucleus. Hence, in spite of the availability of four valence electrons, the crystal has a low conductivity.

The Hole At a very low temperature (say 0°K) the ideal structure of Fig. 2-3 is approached, and the crystal behaves as an insulator, since no free carriers of electricity are available. However, at room temperature, some of the covalent bonds will be broken because of the thermal energy supplied to the crystal, and conduction is made possible. This situation is illustrated in Fig. 2-4. Here an electron, which for the far greater period of time forms part of a covalent bond, is pictured as being dislodged, and therefore free to wander in a random fashion throughout the crystal. The energy E_G required to break such a covalent bond is about 0.72 eV for germanium and 1.1 eV for silicon at room temperature. The absence of the electron in the covalent bond is represented by the small circle in Fig. 2-4, and such an incomplete covalent bond is called a *hole*. The importance of the hole is that it may serve as a carrier of electricity comparable in effectiveness with the free electron.

The mechanism by which a hole contributes to the conductivity is qualitatively as follows: When a bond is incomplete so that a hole exists, it is relatively easy for a valence electron in a neighboring atom to leave its covalent bond to fill this hole. An electron moving from a bond to fill a hole leaves a hole in its initial position. Hence the hole effectively moves in the direction opposite to that of the electron. This hole, in its new position, may now be filled by an electron from another covalent bond, and the hole will correspondingly move one more step in the direction opposite to the motion of the electron. Here we have a mechanism for the conduction of electricity which does not involve *free* electrons. This phenomenon is illustrated schematically in Fig. 2-5, where a circle with a dot in it represents a completed bond, and an empty circle designates a hole. Figure 2-5*a* shows a row of 10 ions, with a broken bond, or hole, at ion 6. Now imagine that an electron from ion 7 moves into the hole at ion 6, so that the configuration of Fig. 2-5*b* results. If we compare this figure with Fig. 2-5*a*, it looks as if the hole in (*a*) has

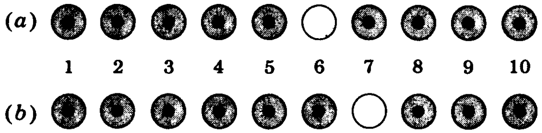


Fig. 2-5 The mechanism by which a hole contributes to the conductivity.

moved toward the right in (*b*) (from ion 6 to ion 7). This discussion indicates that the motion of the hole in one direction actually means the transport of a negative charge an equal distance in the opposite direction. So far as the flow of electric current is concerned, the hole behaves like a positive charge equal in magnitude to the electronic charge. We can consider that the holes are physical entities whose movement constitutes a flow of current. The heuristic argument that a hole behaves as a *free* positive charge carrier may be justified by quantum mechanics.[1] An experimental verification of this concept is given in Sec. 2-6.

In a pure (*intrinsic*) semiconductor the number of holes is equal to the number of free electrons. Thermal agitation continues to produce new hole-electron pairs, whereas other hole-electron pairs disappear as a result of recombination. The hole concentration p must equal the electron concentration n, so that

$$n = p = n_i \tag{2-9}$$

where n_i is called the *intrinsic concentration.*

Effective Mass[2] We digress here briefly to discuss the concept of the effective mass of the electron and hole. It is found that, when quantum mechanics is used to specify the motion within the crystal of a free or conduction electron or hole on which an external field is applied, it is possible to treat the hole and electron as imaginary *classical particles* with effective positive masses m_p and m_n, respectively. This approximation is valid provided that the externally applied fields are much weaker than the internal *periodic* fields produced by the lattice structure. In a perfect crystal these imaginary particles respond only to the external fields.

A wave-mechanical analysis[1] shows that a bound or valence electron *cannot* be treated as a classical particle. This difficulty is bypassed by ignoring the bound electrons and considering only the motion of the holes. In summary, the effective-mass approximation removes the quantum features of the problem and allows us to use Newton's laws to determine the effect of external forces on the (free) electrons and holes within a crystal.

2-3 DONOR AND ACCEPTOR IMPURITIES

If, to intrinsic silicon or germanium, there is added a small percentage of trivalent or pentavalent atoms, a *doped, impure,* or *extrinsic,* semiconductor is formed.

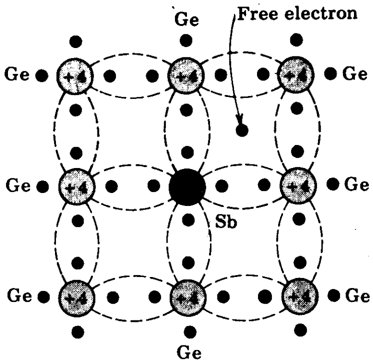


Fig. 2-6 Crystal lattice with a germanium atom displaced by a pentavalent impurity atom.

Donors If the dopant has five valence electrons, the crystal structure of Fig. 2-6 is obtained. The impurity atoms will displace some of the germanium atoms in the crystal lattice. Four of the five valence electrons will occupy covalent bonds, and the fifth will be nominally unbound and will be available as a carrier of current. The energy required to detach this fifth electron from the atom is of the order of only 0.01 eV for Ge or 0.05 eV for Si. Suitable pentavalent impurities are antimony, phosphorus, and arsenic. Such impurities donate excess (negative) electron carriers, and are therefore referred to as *donor*, or *n*-type, impurities.

When donor impurities are added to a semiconductor, allowable energy levels are introduced a very small distance below the conduction band, as is shown in Fig. 2-7. These new allowable levels are essentially a discrete level because the added impurity atoms are far apart in the crystal structure, and hence their interaction is small. In the case of germanium, the distance of the new discrete allowable energy level is only 0.01 eV (0.05 eV in silicon) below the conduction band, and therefore at room temperature almost all the "fifth" electrons of the donor material are raised into the conduction band.

If intrinsic semiconductor material is "doped" with *n*-type impurities, not only does the number of electrons increase, but the number of holes decreases below that which would be available in the intrinsic semiconductor. The reason for the decrease in the number of holes is that the larger number of electrons present increases the rate of recombination of electrons with holes.

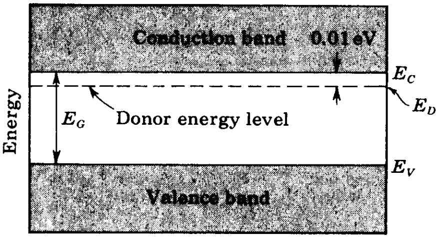


Fig. 2-7 Energy-band diagram of *n*-type semiconductor.

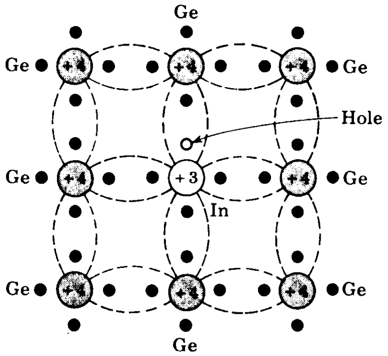


Fig. 2-8 Crystal lattice with a germanium atom displaced by an atom of a trivalent impurity.

Acceptors If a trivalent impurity (boron, gallium, or indium) is added to an intrinsic semiconductor, only three of the covalent bonds can be filled, and the vacancy that exists in the fourth bond constitutes a hole. This situation is illustrated in Fig. 2-8. Such impurities make available positive carriers because they create holes which can accept electrons. These impurities are consequently known as *acceptor*, or *p*-type, impurities. The amount of impurity which must be added to have an appreciable effect on the conductivity is very small. For example, if a donor-type impurity is added to the extent of 1 part in 10^8, the conductivity of germanium at 30°C is multiplied by a factor of 12.

When acceptor, or *p*-type, impurities are added to the intrinsic semiconductor, they produce an allowable discrete energy level which is just above the valence band, as shown in Fig. 2-9. Since a very small amount of energy is required for an electron to leave the valence band and occupy the acceptor energy level, it follows that the holes generated in the valence band by these electrons constitute the largest number of carriers in the semiconductor material.

The Mass-action Law We noted above that adding *n*-type impurities decreases the number of holes. Similarly, doping with *p*-type impurities decreases the concentration of free electrons below that in the intrinsic semiconductor. A theoretical analysis (Sec. 2-12) leads to the result that, under

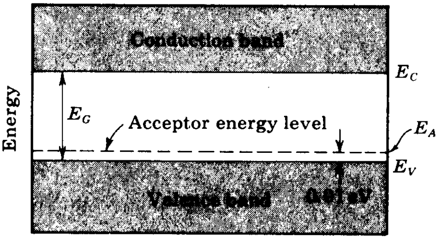


Fig. 2-9 Energy-band diagram of *p*-type semiconductor.

thermal equilibrium, the product of the free negative and positive concentrations is a constant independent of the amount of donor and acceptor impurity doping. This relationship is called the *mass-action law* and is given by

$$np = n_i^2 \tag{2-10}$$

The intrinsic concentration n_i is a function of temperature (Sec. 2-5).

We have the important result that the doping of an intrinsic semiconductor not only increases the conductivity, but also serves to produce a conductor in which the electric carriers are either predominantly holes or predominantly electrons. In an n-type semiconductor, the electrons are called the *majority carriers*, and the holes are called the *minority carriers*. In a p-type material, the holes are the majority carriers, and the electrons are the minority carriers.

2-4 CHARGE DENSITIES IN A SEMICONDUCTOR

Equation (2-10), namely, $np = n_i^2$, gives one relationship between the electron n and the hole p concentrations. These densities are further interrelated by the law of electrical neutrality, which we shall now state in algebraic form: Let N_D equal the concentration of donor atoms. Since, as mentioned above, these are practically all ionized, N_D positive charges per cubic meter are contributed by the donor ions. Hence the total positive-charge density is $N_D + p$. Similarly, if N_A is the concentration of acceptor ions, these contribute N_A negative charges per cubic meter. The total negative-charge density is $N_A + n$. Since the semiconductor is electrically neutral, the magnitude of the positive-charge density must equal that of the negative concentration, or

$$N_D + p = N_A + n \tag{2-11}$$

Consider an n-type material having $N_A = 0$. Since the number of electrons is much greater than the number of holes in an n-type semiconductor ($n \gg p$), then Eq. (2-11) reduces to

$$n \approx N_D \tag{2-12}$$

In an n-type material the free-electron concentration is approximately equal to the density of donor atoms.

In later applications we study the characteristics of n- and p-type materials connected together. Since some confusion may arise as to which type is under consideration at a given moment, we add the subscript n or p for an n-type or a p-type substance, respectively. Thus Eq. (2-12) is more clearly written

$$n_n \approx N_D \tag{2-13}$$

The concentration p_n of holes in the n-type semiconductor is obtained from Eq. (2-10), which is now written $n_n p_n = n_i^2$. Thus

$$p_n = \frac{n_i^2}{N_D} \tag{2-14}$$

Similarly, for a p-type semiconductor,

$$n_p p_p = n_i^2 \qquad p_p \approx N_A \qquad n_p = \frac{n_i^2}{N_A} \tag{2-15}$$

It is possible to add donors to a p-type crystal or, conversely, to add acceptors to n-type material. If equal concentrations of donors and acceptors permeate the semiconductor, it remains intrinsic. The hole of the acceptor combines with the conduction electron of the donor to give no additional free carriers. Thus, from Eq. (2-11) with $N_D = N_A$, we observe that $p = n$, and from Eq. (2-10), $n^2 = n_i^2$, or $n = n_i =$ the intrinsic concentration.

An extension of the above argument indicates that if the concentration of donor atoms added to a p-type semiconductor exceeds the acceptor concentration ($N_D > N_A$), the specimen is changed from a p-type to an n-type semiconductor. [In Eqs. (2-13) and (2-14) N_D should be replaced by $N_D - N_A$.]

2-5 ELECTRICAL PROPERTIES OF GE AND SI

A fundamental difference between a metal and a semiconductor is that the former is *unipolar* [conducts current by means of charges (electrons) of one sign only], whereas a semiconductor is *bipolar* (contains two charge-carrying "particles" of opposite sign).

Conductivity One carrier is negative (the free electron), of mobility μ_n, and the other is positive (the hole), of mobility μ_p. These particles move in opposite directions in an electric field $\mathcal{E}$, but since they are of opposite sign, the current of each is in the same direction. Hence the current density J is given by (Sec. 2-1)

$$J = (n\mu_n + p\mu_p)q\mathcal{E} = \sigma\mathcal{E} \tag{2-16}$$

where n = magnitude of free-electron (negative) concentration
p = magnitude of hole (positive) concentration
σ = conductivity

Hence $\sigma = (n\mu_n + p\mu_p)q$ (2-17)

For the pure semiconductor, $n = p = n_i$, where n_i is the intrinsic concentration.

Intrinsic Concentration With increasing temperature, the density of hole-electron pairs increases and, correspondingly, the conductivity increases. In Sec. 19-5 it is found that the intrinsic concentration n_i varies with T as

$$n_i^2 = A_o T^3 \epsilon^{-E_{GO}/kT} \tag{2-18}$$

where E_{GO} is the energy gap at 0°K in electron volts, k is the Boltzman constant in eV/°K (Appendix A), and A_o is a constant independent of T. The constants E_{GO}, μ_n, μ_p, and many other important physical quantities for germanium and

TABLE 2-1 Properties of germanium and silicon†

Property	Ge	Si
Atomic number	32	14
Atomic weight	72.6	28.1
Density, g/cm^3	5.32	2.33
Dielectric constant (relative)	16	12
Atoms/cm^3	4.4×10^{22}	5.0×10^{22}
E_{GO}, eV, at 0°K	0.785	1.21
E_G, eV, at 300°K	0.72	1.1
n_i at 300°K, cm^{-3}	2.5×10^{13}	1.5×10^{10}
Intrinsic resistivity at 300°K, Ω-cm	45	230,000
μ_n, cm^2/V-s at 300°K	3,800	1,300
μ_p, cm^2/V-s at 300°K	1,800	500
D_n, cm^2/s $= \mu_n V_T$	99	34
D_p, cm^2/s $= \mu_p V_T$	47	13

† G. L. Pearson and W. H. Brattain, History of Semiconductor Research, *Proc. IRE*, vol. 43, pp. 1794–1806, December, 1955. E. M. Conwell, Properties of Silicon and Germanium, Part II, *Proc. IRE*, vol. 46, no. 6, pp. 1281–1299, June, 1958.

silicon are given in Table 2-1. Note that germanium has of the order of 10^{22} atoms/cm^3, whereas at room temperature (300°K), $n_i \approx 10^{13}/\text{cm}^3$. Hence only 1 atom in about 10^9 contributes a free electron (and also a hole) to the crystal because of broken covalent bonds. For silicon this ratio is even smaller, about 1 atom in 10^{12}

The Energy Gap The forbidden region E_G in a semiconductor depends upon temperature, as pointed out in Sec. 1-7. Experimentally it is found that, for silicon,[3]

$$E_G(T) = 1.21 - 3.60 \times 10^{-4}T \tag{2-19}$$

and at room temperature (300°K), $E_G = 1.1$ eV. Similarly, for germanium,[4]

$$E_G(T) = 0.785 - 2.23 \times 10^{-4}T \tag{2-20}$$

and at room temperature, $E_G = 0.72$ eV.

The Mobility This parameter μ varies[3] as T^{-m} over a temperature range of 100 to 400°K. For silicon, $m = 2.5$ (2.7) for electrons (holes), and for germanium, $m = 1.66$ (2.33) for electrons (holes). The mobility is also found[4] to be a function of electric field intensity and remains constant only if $\mathcal{E} < 10^3$ V/cm in n-type silicon. For $10^3 < \mathcal{E} < 10^4$ V/cm, μ_n varies approximately as $\mathcal{E}^{-\frac{1}{2}}$. For higher fields, μ_n is inversely proportional to $\mathcal{E}$ and the carrier speed approaches the constant value of 10^7 cm/s.

EXAMPLE (*a*) Using Avogadro's number, verify the numerical value given in Table 2-1 for the concentration of atoms in germanium. (*b*) Find the resistivity of intrinsic germanium at 300°K. (*c*) If a donor-type impurity is added to the extent of 1 part in 10^8 germanium atoms, find the resistivity. (*d*) If germanium were a monovalent metal, find the ratio of its conductivity to that of the *n*-type semiconductor in part *c*.

Solution *a*. A quantity of any substance equal to its molecular weight in grams is a *mole* of that substance. Further, a mole of any substance contains the same number of molecules as a mole of any other material. This number is called *Avogadro's number* and equals 6.02×10^{23} molecules per mole (Appendix A). Thus, for monatomic germanium (using Table 2-1),

$$\text{Concentration} = 6.02 \times 10^{23} \frac{\text{atoms}}{\text{mole}} \times \frac{1 \text{ mole}}{72.6 \text{ g}} \times \frac{5.32 \text{ g}}{\text{cm}^3} = 4.41 \times 10^{22} \frac{\text{atoms}}{\text{cm}^3}$$

b. From Eq. (2-17), with $n = p = n_i$,

$$\sigma = n_i q(\mu_n + \mu_p) = (2.5 \times 10^{13}\ \text{cm}^{-3})(1.60 \times 10^{-19}\ \text{C})(3{,}800 + 1{,}800)\ \frac{\text{cm}^2}{\text{V-s}}$$
$$= 0.0224\ (\Omega\text{-cm})^{-1}$$

$$\text{Resistivity} = \frac{1}{\sigma} = \frac{1}{0.0224} = 44.6\ \Omega\text{-cm}$$

in agreement with the value in Table 2-1.

c. If there is 1 donor atom per 10^8 germanium atoms, then $N_D = 4.41 \times 10^{14}$ atoms/cm³. From Eq. (2-12) $n \approx N_D$ and from Eq. (2-14)

$$p = \frac{n_i^2}{N_D} = \frac{(2.5 \times 10^{13})^2}{4.41 \times 10^{14}} = 1.42 \times 10^{12}\ \text{holes/cm}^3$$

Since $n \gg p$, we can neglect p in calculating the conductivity. From Eq. (2-17)

$$\sigma = nq\mu_n = 4.41 \times 10^{14} \times 1.60 \times 10^{-19} \times 3{,}800 = 0.268\ (\Omega\text{-cm})^{-1}$$

The resistivity $= 1/\sigma = 1/0.268 = 3.72\ \Omega$-cm.

NOTE: The addition of 1 donor atom in 10^8 germanium atoms has multiplied the conductivity by a factor of $44.6/3.72 = 11.7$.

d. If each atom contributed one free electron to the "metal," then

$$n = 4.41 \times 10^{22}\ \text{electrons/cm}^3$$

and

$$\sigma = nq\mu_n = 4.41 \times 10^{22} \times 1.60 \times 10^{-19} \times 3{,}800$$
$$= 2.58 \times 10^{7}\ (\Omega\text{-cm})^{-1}$$

Hence the conductivity of the "metal" is higher than that of the *n*-type semiconductor by a factor of

$$\frac{2.58 \times 10^7}{0.268} \approx 10^8$$

2-6 THE HALL EFFECT

If a specimen (metal or semiconductor) carrying a current **I** is placed in a transverse magnetic field **B**, an electric field $\mathcal{E}$ is induced in the direction perpendicular to both **I** and **B**. This phenomenon, known as the *Hall effect*, is used to determine whether a semiconductor is n- or p-type and to find the carrier concentration. Also, by simultaneously measuring the conductivity σ, the mobility μ can be calculated.

The physical origin of the Hall effect is not difficult to find. If in Fig. 2-10 **I** is in the positive X direction and **B** is in the positive Z direction, a force will be exerted in the negative Y direction on the current carriers. The current **I** may be due to holes moving from left to right or to free electrons traveling from right to left in the semiconductor specimen. Hence, independently of whether the carriers are holes or electrons, they will be forced downward toward side 1 in Fig. 2-10. If the semiconductor is n-type material, so that the current is carried by electrons, these electrons will accumulate on side 1, and this surface becomes negatively charged with respect to side 2. Hence a potential, called the *Hall voltage*, appears between surfaces 1 and 2.

If the polarity of V_H is positive at terminal 2, then, as explained above, the carriers must be electrons. If, on the other hand, terminal 1 becomes charged positively with respect to terminal 2, the semiconductor must be p-type. These results have been verified experimentally, thus justifying the bipolar (two-carrier) nature of the current in a semiconductor.

If **I** is the current in a p-type semiconductor, the carriers might be considered to be the *bound* electrons jumping from right to left. Then side 1 would become negatively charged. However, experimentally, side 1 is found to become positive with respect to side 2 for a p-type specimen. This experiment confirms the quantum-mechanical fact noted in Sec. 2-2 that the hole acts like a classical *free* positive-charge carrier.

Experimental Determination of Mobility In the equilibrium state the electric field intensity $\mathcal{E}$ due to the Hall effect must exert a force on the carrier which just balances the magnetic force, or

$$q\mathcal{E} = Bqv \tag{2-21}$$

where q is the magnitude of the charge on the carrier, and v is the drift speed. From Eq. (1-3), $\mathcal{E} = V_H/d$, where d is the distance between surfaces 1 and 2.

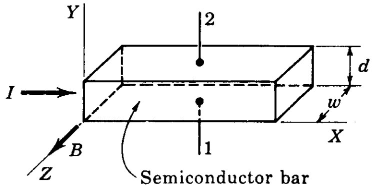


Fig. 2-10 Pertaining to the Hall effect. The carriers (whether electrons or holes) are subjected to a magnetic force in the negative Y direction.

From Eq. (2-6), $J = \rho v = I/wd$, where J is the current density, ρ is the charge density, and w is the width of the specimen in the direction of the magnetic field. Combining these relationships, we find

$$V_H = \mathcal{E}d = Bvd = \frac{BJd}{\rho} = \frac{BI}{\rho w} \tag{2-22}$$

If V_H, B, I, and w are measured, the charge density ρ can be determined from Eq. (2-22).

It is customary to introduce the Hall coefficient R_H defined by

$$R_H \equiv \frac{1}{\rho} \tag{2-23}$$

Hence $$R_H = \frac{V_H w}{BI} \tag{2-24}$$

If conduction is due primarily to charges of one sign, the conductivity σ is related to the mobility μ by Eq. (2-8), or

$$\sigma = \rho\mu \tag{2-25}$$

If the conductivity is measured together with the Hall coefficient, the mobility can be determined from

$$\mu = \sigma R_H \tag{2-26}$$

We have assumed in the foregoing discussion that all particles travel with the mean drift speed v. Actually, the current carriers have a random thermal distribution in speed. If this distribution is taken into account, it is found that Eq. (2-24) remains valid provided that R_H is defined by $3\pi/8\rho$. Also, Eq. (2-26) must be modified to $\mu = (8\sigma/3\pi)R_H$.

Applications Since V_H is proportional to B (for a given current I), then the Hall effect has been incorporated into a magnetic field meter. Another instrument, called a *Hall-effect multiplier*, is available to give an output proportional to the product of two signals. If I is made proportional to one of the inputs and if B is linearly related to the second signal, then, from Eq. (2-22), V_H is proportional to the product of the two inputs.

2-7 CONDUCTIVITY MODULATION

Since the conductivity σ of a semiconductor is proportional to the concentration of free carriers [Eq. (2-17)], σ may be increased by increasing n or p. The two most important methods for varying n and p are to change the temperature or to illuminate the semiconductor and thereby generate new hole-electron pairs.

Thermistors The conductivity of germanium (silicon) is found from Eq. (2-18) to increase approximately 6 (8) percent per degree increase in temperature. Such a large change in conductivity with temperature places a limitation upon the use of semiconductor devices in some circuits. On the other hand, for some applications it is exactly this property of semiconductors that is used to advantage. A semiconductor used in this manner is called a *thermistor.* Such a device finds extensive application in thermometry, in the measurement of microwave-frequency power, as a thermal relay, and in control devices actuated by changes in temperature. Silicon and germanium are not used as thermistors because their properties are too sensitive to impurities. Commercial thermistors consist of sintered mixtures of such oxides as NiO, Mn_2O_3, and Co_2O_3.

The exponential decrease in resistivity (reciprocal of conductivity) of a semiconductor should be contrasted with the small and almost linear increase in resistivity of a metal. An increase in the temperature of a metal results in greater thermal motion of the ions, and hence decreases slightly the mean free path of the free electrons. The result is a decrease in the mobility, and hence in conductivity. For most metals the resistance increases about 0.4 percent/°C increase in temperature. It should be noted that a thermistor has a negative coefficient of resistance, whereas that of a metal is positive and of much smaller magnitude. By including a thermistor in a circuit it is possible to compensate for temperature changes over a range as wide as 100°C.

A heavily doped semiconductor can exhibit a positive temperature coefficient of resistance, for under these circumstances the material acquires metallic properties and the resistance increases because of the decrease in carrier mobility with temperature. Such a device, called a *sensistor* (manufactured by Texas Instruments), has a temperature coefficient of resistance of +0.7 percent/°C (over the range from −60 to +150°C).

Photoconductors If radiation falls upon a semiconductor, its conductivity increases. This photoconductive effect is explained as follows: Radiant energy supplied to the semiconductor ionizes covalent bonds; that is, these bonds are broken, and hole-electron pairs in excess of those generated thermally are created. These increased current carriers decrease the resistance of the material, and hence such a device is called a *photoresistor,* or *photoconductor.* For a light-intensity change of 100 fc,† the resistance of a commercial photoconductor may change by several kilohms.

In Fig. 2-11 we show the energy diagram of a semiconductor having both acceptor and donor impurities. If photons of sufficient energies illuminate this specimen, *photogeneration* takes place and the following transitions are possible: An electron-hole pair can be created by a high-energy photon, in what is called *intrinsic excitation;* a photon may excite a donor electron into the conduction band; or a valence electron may go into an acceptor state. The

---

† fc is the standard abbreviation for foot-candle.

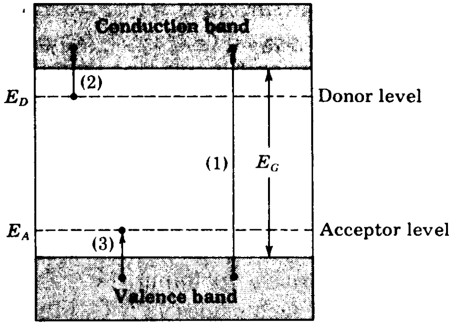


Fig. 2-11 Photoexcitation in semiconductors. (1) is intrinsic whereas (2) and (3) are extrinsic excitations.

last two transitions are known as *impurity excitations.* Since the density of states in the conduction and valence bands greatly exceeds the density of impurity states, photoconductivity is due principally to intrinsic excitation.

Spectral Response The minimum energy of a photon required for intrinsic excitation is the forbidden-gap energy E_G (electron volts) of the semiconductor material. The wavelength λ_c of a photon whose energy corresponds to E_G is given by Eq. (1-14), with $E_1 - E_2 = E_G$. If λ_c is expressed in microns† and E_G in electron volts,

$$\lambda_c = \frac{1.24}{E_G} \tag{2-27}$$

If the wavelength λ of the radiation exceeds λ_c, then the energy of the photon is less than E_G and such a photon cannot cause a valence electron to enter the conduction band. Hence λ_c is called the *critical,* or *cutoff, wavelength,* or *long-wavelength threshold,* of the material. For Si, $E_G = 1.1$ eV and $\lambda_c = 1.13\ \mu$m, whereas for Ge, $E_G = 0.72$ eV and $\lambda_c = 1.73\ \mu$m at room temperature (Table 2-1).

The spectral-sensitivity curves for Si and Ge are plotted in Fig. 2-12, indicating that a photoconductor is a frequency-selective device. This means

---

† 1 micron = 1 micrometer = 1 μm $\doteq 10^{-6}$ m.

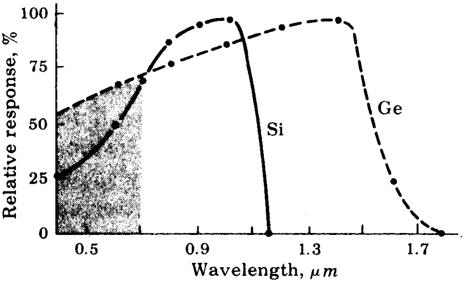


Fig. 2-12 Relative spectral response of Si and Ge. (Courtesy of Texas Instruments, Inc.)

that a given intensity of light of one wavelength will not generate the same number of free carriers as an equal intensity of light of another wavelength. In other words, the *photoelectric yield,* or *spectral response,* depends upon the frequency of the incident radiation. Note that the long-wavelength limit is slightly greater than the values of λ_c calculated above, because of the impurity excitations. As the wavelength is decreased ($\lambda < \lambda_c$ or $f > f_c$), the response increases and reaches a maximum. The range of wavelengths of visible light (0.38 to 0.76 μm) is indicated by the shaded region in Fig. 2-12.

Commercial Photoconductive Cells There are three important types of applications of such a device: It is used (1) to measure a fixed amount of illumination (as with light meter), (2) to record a modulating light intensity (as on a sound track), and (3) as an ON–OFF light relay (as in a digital or control circuit).

The photoconducting device with the widest application is the cadmium sulfide cell. The sensitive area of this device consists of a layer of chemically deposited CdS, which may contain a small amount of silver, antimony, or indium impurities. In absolute darkness the resistance may be as high as 2 M, and when stimulated with strong light, the resistance may be less than 10 Ω.

The primary advantages of CdS photoconductors are their high dissipation capability, their excellent sensitivity in the visible spectrum, and their low resistance when stimulated by light. These photoconductors are designed to dissipate safely 300 mW, and can be made to handle safely power levels of several watts. Hence a CdS photoconductor can operate a relay directly, without intermediate amplifier circuits.

Other types of photoconductive devices are available for specific applications. A lead sulfide, PbS, cell has a peak on the sensitivity curve at 2.9 μm, and hence is used for infrared-detection or infrared-absorption measurements. A selenium cell is sensitive throughout the visible end, and particularly toward the blue end of the spectrum.

2-8 GENERATION AND RECOMBINATION OF CHARGES

In Sec. 2-2 we see that in a pure semiconductor the number of holes is equal to the number of free electrons. Thermal agitation, however, continues to generate g new hole-electron pairs per unit volume per second, while other hole-electron pairs disappear as a result of recombination; in other words, free electrons fall into empty covalent bonds, resulting in the loss of a pair of mobile carriers. On an average, a hole (an electron) will exist for $\tau_p(\tau_n)$ s before recombination. This time is called the *mean lifetime* of the hole and electron, respectively. These parameters are very important in semiconductor devices because they indicate the time required for electron and hole concentrations which have been caused to change to return to their equilibrium concentrations.

Consider a bar of n-type silicon containing the thermal-equilibrium con-

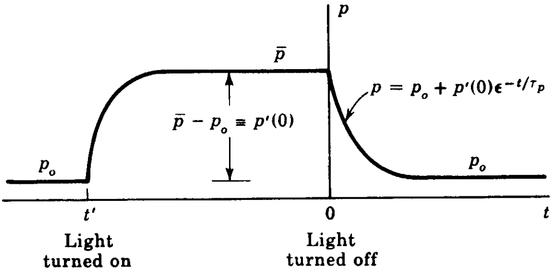


Fig. 2-13 The hole (minority) concentration in an n-type semiconductor bar as a function of time, due to generation and recombination.

centration p_o and n_o. Assume that at $t = t'$ the specimen is illuminated (Fig. 2-13) and that additional hole-electron pairs are generated uniformly throughout the crystal. An equilibrium situation is reached, and the new concentrations are $\bar{p}$ and $\bar{n}$ under the influence of the radiation. The *photo-injected*, or *excess*, concentration is $\bar{p} - p_o$ for holes and $\bar{n} - n_o$ for electrons. Since the radiation causes hole-electron pairs to be created, then clearly,

$$\bar{p} - p_o = \bar{n} - n_o \qquad (2\text{-}28)$$

Although the increase in hole concentration p equals that for the electron density n, the *percentage* increase for electrons in an n-type semiconductor (where electrons are plentiful) is very small. On the other hand, the percentage increase in holes may be tremendous, because holes are scarce in an n-type crystal. In summary, the radiation affects the majority concentration hardly at all, and therefore we shall limit the discussion to the behavior of the minority carriers.

After a steady state is reached, at $t = 0$ in Fig. 2-13, the radiation is removed. We shall now demonstrate that the excess carrier density returns to zero exponentially with time. To do so we must derive the differential equation which governs the hole concentration as a function of time for $t > 0$ (*when there is no external excitation*).

From the definition of mean lifetime τ_p and *assuming that τ_p is independent of the magnitude of the hole concentration*,

$$\frac{p}{\tau_p} = \textit{decrease}\text{ in hole concentration per second due to recombination} \qquad (2\text{-}29)$$

From the definition of the generation rate,

$$g = \textit{increase}\text{ in hole concentration per second due to thermal generation} \qquad (2\text{-}30)$$

Since charge can neither be created nor destroyed, there must be an increase in hole concentration per second of amount dp/dt. This rate must, at every instant of time, equal the algebraic sum of the rates given in Eqs. (2-29) and

(2-30), or

$$\frac{dp}{dt} = g - \frac{p}{\tau_p} \tag{2-31}$$

Under steady-state conditions, $dp/dt = 0$, and with no radiation falling on the sample, the hole concentration p reaches its thermal-equilibrium value p_o. Hence $g = p_o/\tau_p$, and the above equation becomes

$$\frac{dp}{dt} = \frac{p_o - p}{\tau_p} \tag{2-32}$$

The *excess*, or *injected*, carrier density p' is defined as the increase in minority concentration above the equilibrium value. Since p' is a function of time, then

$$p' \equiv p - p_o = p'(t) \tag{2-33}$$

It follows from Eq. (2-32) that the differential equation controlling p' is

$$\frac{dp'}{dt} = -\frac{p'}{\tau_p} \tag{2-34}$$

The rate of change of excess concentration is proportional to this concentration—an intuitively correct result. The minus sign indicates that the change is a decrease in the case of recombination and an increase when the concentration is recovering from a temporary depletion.

Since the radiation results in an initial (at $t \leq 0$) excess concentration $p'(0) = \bar{p} - p_o$ and then this excitation is removed, the solution of Eq. (2-34) for $t \geq 0$ is

$$p'(t) = p'(0)\epsilon^{-t/\tau_p} = (\bar{p} - p_o)\epsilon^{-t/\tau_p} = p - p_o \tag{2-35}$$

The excess concentration decreases exponentially to zero ($p' = 0$ or $p = p_o$) with a time constant equal to the mean lifetime τ_p, as indicated in Fig. 2-13. The pulsed-light method indicated in this figure is used to measure τ_p.

Recombination Centers Recombination is the process where an electron moves from the conduction band into the valence band so that a mobile electron-hole pair disappear. Classical mechanics requires that momentum be conserved in an encounter of two particles. Since the momentum is zero after recombination, this conservation law requires that the "colliding" electron and hole must have equal magnitudes of momentum and they must be traveling in opposite directions. This requirement is very stringent, and hence the probability of recombination by such a direct encounter is very small.

The most important mechanism in silicon or germanium through which holes and electrons recombine is that involving *traps*, or *recombination centers*,[5] which contribute electronic states in the energy gap of the semiconductor. Such a location acts effectively as a third body which can satisfy the conservation-of-momentum requirement. These new states are associated with imperfections in the crystal. Specifically, metallic impurities in the semiconductor are capable of introducing energy states in the forbidden gap. Recombination is affected not only by volume impurities but also by surface imperfections in the crystal.

Gold is extensively used as a recombination agent by semiconductor-device manufacturers. Thus the device designer can obtain desired carrier lifetimes by introducing gold into silicon under controlled conditions.[6] Carrier lifetimes range from nanoseconds (1 ns = 10^{-9} s) to hundreds of microseconds (μs).

2-9 DIFFUSION

In addition to a conduction current, the transport of charges in a semiconductor may be accounted for by a mechanism called *diffusion*, not ordinarily encountered in metals. The essential features of diffusion are now discussed.

It is possible to have a nonuniform concentration of particles in a semiconductor. As indicated in Fig. 2-14, the concentration p of holes varies with distance x in the semiconductor, and there exists a concentration gradient, dp/dx, in the density of carriers. The existence of a gradient implies that if an imaginary surface (shown dashed) is drawn in the semiconductor, the density of holes immediately on one side of the surface is larger than the density on the other side. The holes are in a random motion as a result of their thermal energy. Accordingly, holes will continue to move back and forth across this surface. We may then expect that, in a given time interval, more holes will cross the surface from the side of greater concentration to the side of smaller concentration than in the reverse direction. This net transport of holes across the surface constitutes a current in the positive X direction. It should be noted that this net transport of charge is not the result of mutual repulsion among charges of like sign, but is simply the result of a statistical phenomenon. This diffusion is exactly analogous to that which occurs in a neutral gas if a concentration gradient exists in the gaseous container. The diffusion hole-current density J_p (amperes per square meter) is proportional

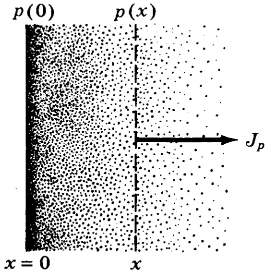


Fig. 2-14 A nonuniform concentration $p(x)$ results in a diffusion current J_p.

to the concentration gradient, and is given by

$$J_p = -qD_p \frac{dp}{dx} \tag{2-36}$$

where D_p (square meters per second) is called the *diffusion constant* for holes. Since p in Fig. 2-14 decreases with increasing x, then dp/dx is negative and the minus sign in Eq. (2-36) is needed, so that J_p will be positive in the positive X direction. A similar equation exists for diffusion electron-current density [p is replaced by n, and the minus sign is replaced by a plus sign in Eq. (2-36)].

Einstein Relationship Since both diffusion and mobility are statistical thermodynamic phenomena, D and μ are not independent. The relationship between them is given by the Einstein equation (Eq. 19-59)

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = V_T \tag{2-37}$$

where V_T is the "volt-equivalent of temperature," defined by

$$V_T \equiv \frac{\bar{k}T}{q} = \frac{T}{11{,}600} \tag{2-38}$$

where $\bar{k}$ is the Boltzmann constant in joules per degree Kelvin. Note the distinction between $\bar{k}$ and k; the latter is the Boltzmann constant in electron volts per degree Kelvin. (Numerical values of $\bar{k}$ and k are given in Appendix A. From Sec. 1-3 it follows that $\bar{k} = 1.60 \times 10^{-19}k$.) At room temperature (300°K), $V_T = 0.026$ V, and $\mu = 39D$. Measured values of μ and computed values of D for silicon and germanium are given in Table 2-1.

Total Current It is possible for both a potential gradient and a concentration gradient to exist simultaneously within a semiconductor. In such a situation the total hole current is the sum of the drift current [Eq. (2-7), with n replaced by p] and the diffusion current [Eq. (2-36)], or

$$J_p = q\mu_p p\mathcal{E} - qD_p \frac{dp}{dx} \tag{2-39}$$

Similarly, the net electron current is

$$J_n = q\mu_n n\mathcal{E} + qD_n \frac{dn}{dx} \tag{2-40}$$

2-10 THE CONTINUITY EQUATION

In Sec. 2-8 it was seen that if we disturb the equilibrium concentrations of carriers in a semiconductor material, the concentration of holes or electrons (which is constant throughout the crystal) will vary with time. In the general

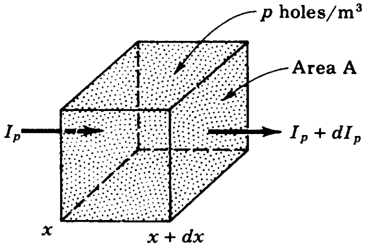


Fig. 2-15 Relating to the conservation of charge.

case, however, the carrier concentration in the body of a semiconductor is a function of both time and distance. We now derive the differential equation which governs this functional relationship. This equation is based on the fact that charge can be neither created nor destroyed, and hence is an extension of Eq. (2-31).

Consider the infinitesimal element of volume of area A and length dx (Fig. 2-15) within which the average hole concentration is p. Assume that the problem is one-dimensional and that the hole current I_p is a function of x. If, as indicated in Fig. 2-15, the current entering the volume at x is I_p at time t and leaving at $x + dx$ is $I_p + dI_p$ at the same time t, there must be dI_p more coulombs per second leaving the volume than entering it (for a positive value of dI_p). Hence the *decrease* in number of coulombs per second within the volume is dI_p. Since the magnitude of the carrier charge is q, then dI_p/q equals the decrease in the number of holes per second within the elemental volume $A\,dx$. Remembering that the current density $J_p = I_p/A$, we have

$$\frac{1}{qA}\frac{dI_p}{dx} = \frac{1}{q}\frac{dJ_p}{dx} = \textit{decrease}\text{ in hole concentration (holes per unit volume) per second, due to current } I_p \qquad (2\text{-}41)$$

From Eq. (2-30) we know that there is an *increase* per second of $g = p_o/\tau_p$ holes per unit volume due to thermal generation, and from Eq. (2-29) a *decrease* per second of p/τ_p holes per unit volume because of recombination. Since charge can neither be created nor destroyed, the *increase* in holes per unit volume per second, dp/dt, must equal the algebraic sum of all the increases listed above, or

$$\frac{\partial p}{\partial t} = \frac{p_o - p}{\tau_p} - \frac{1}{q}\frac{\partial J_p}{\partial x} \qquad (2\text{-}42)$$

(Since both p and J_p are functions of both t and x, then partial derivatives are used in this equation.)

The continuity equation is applied to a specific physical problem in the following section and is discussed further in Sec. 19-9. Equation (2-42) is called the *law of conservation of charge*, or the *continuity equation* for charge. This law applies equally well for electrons, and the corresponding equation is obtained by replacing p by n in Eq. (2-42).

2-11 INJECTED MINORITY-CARRIER CHARGE

Consider the physical situation pictured[7] in Fig. 2-16*a*. A long semiconductor bar is doped uniformly with donor atoms so that the concentration $n = N_D$ is independent of position. Radiation falls upon the end of the bar at $x = 0$. Some of the photons are captured by the bound electrons in the covalent bonds near the illuminated surface. As a result of this energy transfer, these bonds are broken and hole-electron *pairs* are generated. Let us investigate how the steady-state minority-carrier concentration p varies with the distance x into the specimen.

We shall make the reasonable assumption that the injected minority concentration is very small compared with the doping level; that is, $p' < n$. The statement that the minority concentration is much smaller than the majority concentration is called the *low-level injection* condition. Since the drift current is proportional to the concentration [Eq. (2-16)] and since $p = p' + p_o \ll n$, we shall neglect the *hole* drift current (but not the *electron* drift current) and shall assume that I_p is due entirely to diffusion. This assumption is justified at the end of this section. The controlling differential equation for p is

$$\frac{d^2p}{dx^2} = \frac{p - p_o}{D_p\tau_p} \tag{2-43}$$

This equation is obtained by substituting Eq. (2-36) for the diffusion current into the equation of continuity [Eq. (2-42)] and setting $dp/dt = 0$ for steady-state operation. Defining the *diffusion length* for holes L_p by

$$L_p \equiv (D_p\tau_p)^{\frac{1}{2}} \tag{2-44}$$

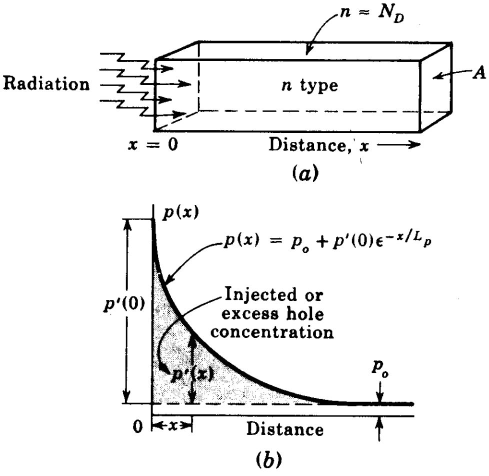


Fig. 2-16 (*a*) Light falls upon the end of a long semiconductor bar. This excitation causes hole-electron pairs to be injected at $x = 0$. (*b*) The hole (minority) concentration $p(x)$ in the bar as a function of distance x from the end of the specimen. The injected concentration is $p'(x) = p(x) - p_o$. The radiation injects $p'(0)$ carriers/m³ into the bar at $x = 0$. [Not drawn to scale since $p'(0) \gg p_o$.]

the differential equation for the injected hole concentration $p' = p - p_o$ becomes

$$\frac{d^2p'}{dx^2} = \frac{p'}{L_p^2} \tag{2-45}$$

The solution of this equation is

$$p'(x) = K_1\epsilon^{-x/L_p} + K_2\epsilon^{+x/L_p} \tag{2-46}$$

where K_1 and K_2 are constants of integration. Consider a very long piece of semiconductor extending from $x = 0$ in the positive X direction. Since the concentration cannot become infinite as $x \to \infty$, then K_2 must be zero. We shall assume that at $x = 0$ the injected concentration is $p'(0)$. To satisfy this boundary condition, $K_1 = p'(0)$. Hence

$$p'(x) = p'(0)\epsilon^{-x/L_p} = p(x) - p_o \tag{2-47}$$

The hole concentration decreases exponentially with distance, as indicated in Fig. 2-16*b*. We see that the diffusion length L_p represents the distance into the semiconductor at which the injected concentration falls to $1/\epsilon$ of its value at $x = 0$. In Sec. 19-9 it is demonstrated that L_p also represents the average distance that an injected hole travels before recombining with an electron.

Diffusion Currents The minority (hole) diffusion current is $I_p = AJ_p$, where A is the cross section of the bar. From Eqs. (2-36) and (2-47)

$$I_p(x) = \frac{AqD_pp'(0)}{L_p}\epsilon^{-x/L_p} = \frac{AqD_p}{L_p}[p(0) - p_o]\epsilon^{-x/L_p} \tag{2-48}$$

This current falls exponentially with distance in the same manner that the minority-carrier concentration decreases. This result is used to find the current in a semiconductor diode (Sec. 3-3).

The majority (electron) diffusion current is $AqD_n\, dn/dx$. Assuming that electrical neutrality is preserved under low-level injection, then $n' = p'$, or

$$n - n_o = p - p_o \tag{2-49}$$

Since the thermal-equilibrium concentrations n_o and p_o are independent of the position x, then

$$\frac{dn}{dx} = \frac{dp}{dx} \tag{2-50}$$

Hence the electron diffusion current is

$$AqD_n\frac{dn}{dx} = AqD_n\frac{dp}{dx} = -\frac{D_n}{D_p}I_p \tag{2-51}$$

where $I_p = -AqD_p\, dp/dx$ = the hole diffusion current. The dependence of the diffusion current upon x is given in Eq. (2-48). The magnitude of the ratio of majority to minority diffusion current is $D_n/D_p \sim 2$ for germanium and ~ 3 for silicon (Table 2-1).

Drift Currents Since Fig. 2-16*a* represents an open-circuited bar, the resultant current (the sum of hole and electron currents) must be zero everywhere. Hence a majority (electron) drift current I_{nd} must exist such that

$$I_p + \left(I_{nd} - \frac{D_n I_p}{D_p}\right) = 0 \tag{2-52}$$

or

$$I_{nd} = \left(\frac{D_n}{D_p} - 1\right) I_p \tag{2-53}$$

From Eq. (2-48) we see that the electron drift current also decreases exponentially with distance.

It is important to point out that *an electric field* $\mathcal{E}$ *must exist in the bar* in order for a drift current to exist. This field is created internally by the injected carriers. From Eqs. (2-7) and (2-53)

$$\mathcal{E} = \frac{1}{Aqn\mu_n}\left(\frac{D_n}{D_p} - 1\right) I_p \tag{2-54}$$

The results obtained in this section are based on the assumption that the hole drift current I_{pd} is zero. Using Eq. (2-7), with n replaced by p, the next approximation for this current is

$$I_{pd} = Aqp\mu_p\mathcal{E} = \frac{p}{n}\frac{\mu_p}{\mu_n}\left(\frac{D_n}{D_p} - 1\right) I_p \tag{2-55}$$

Since $p \ll n$, then $I_{pd} \ll I_p$. The hole drift current is negligible compared with the hole diffusion current, thus justifying the assumption that the injected minority-carrier current, under low-level injection, is essentially a diffusion current.

2-12 THE POTENTIAL VARIATION WITHIN A GRADED SEMICONDUCTOR

Consider a semiconductor (Fig. 2-17*a*) where the hole concentration p is a function of x; that is, the doping is *nonuniform*, or graded.[7] Assume a steady-state situation and zero excitation; that is, no carriers are injected into the specimen from any *external* source. With no excitation there can be no *steady* movement of charge in the bar, although the carriers possess random motion due to thermal agitation. Hence the total hole current must be zero. (Also, the total electron current must be zero.) Since p is not constant, we expect a nonzero hole diffusion current. In order for the total hole current to vanish there must exist a hole drift current which is equal and opposite to the diffusion current. However, a conduction current requires an electric field,

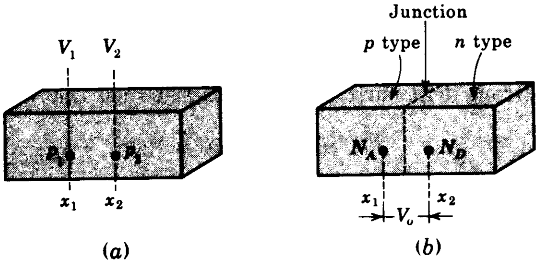


Fig. 2-17 (a) A graded semiconductor: $p(x)$ is not constant. (b) One portion is doped uniformly with acceptor ions and the other section is doped uniformly with donor ions so that a metallurgical junction is formed. A contact potential V_o appears across this step-graded p-n junction.

and hence we conclude that, as a result of the nonuniform doping, an electric field is generated within the semiconductor. We shall now find this field and the corresponding potential variation throughout the bar.

Setting $J_p = 0$ in Eq. (2-39) and using the Einstein relationship $D_p = \mu_p V_T$ [Eq. (2-37)], we obtain

$$\mathcal{E} = \frac{V_T}{p}\frac{dp}{dx} \tag{2-56}$$

If the doping concentration $p(x)$ is known, this equation allows the built-in field $\mathcal{E}(x)$ to be calculated. From $\mathcal{E} = -dV/dx$ we can calculate the potential variation. Thus

$$dV = -V_T \frac{dp}{p} \tag{2-57}$$

If this equation is integrated between x_1, where the concentration is p_1 and the potential is V_1 (Fig. 2-17a), and x_2, where $p = p_2$ and $V = V_2$, the result is

$$V_{21} \equiv V_2 - V_1 = V_T \ln \frac{p_1}{p_2} \tag{2-58}$$

Note that the potential difference between two points depends only upon the concentrations at these two points and is independent of their separation $x_2 - x_1$. Equation (2-58) may be put in the form

$$p_1 = p_2 \epsilon^{V_{21}/V_T} \tag{2-59}$$

This is the Boltzmann relationship of kinetic gas theory.

Mass-action Law Starting with $J_n = 0$ and proceeding as above, the Boltzmann equation for electrons is obtained.

$$n_1 = n_2 \epsilon^{-V_{21}/V_T} \tag{2-60}$$

Multiplying Eqs. (2-59) and (2-60) gives

$$n_1 p_1 = n_2 p_2 \tag{2-61}$$

This equation states that the product np is a constant independent of x, and hence of the amount of doping, under thermal equilibrium. For an intrinsic semiconductor, $n = p = n_i$, and hence

$$np = n_i^2 \tag{2-10}$$

which is the law of mass action introduced in Sec. 2-3. An alternative proof is given in Sec. 19-5.

An Open-circuited Step-graded Junction Consider the special case indicated in Fig. 2-17*b*. The left half of the bar is p-type with a constant concentration N_A, whereas the right-half is n-type with a uniform density N_D. The dashed plane is a metallurgical (p-n) junction separating the two sections with different concentration. This type of doping, where the density changes abruptly from p- to n-type, is called *step grading*. The step-graded junction is located at the plane where the concentration is zero. The above theory indicates that there is a built-in potential between these two sections (called the *contact difference of potential* V_o). Equation (2-58) allows us to calculate V_o. Thus

$$V_o = V_{21} = V_T \ln \frac{p_{po}}{p_{no}} \tag{2-62}$$

because $p_1 = p_{po}$ = thermal-equilibrium hole concentration in p side and $p_2 = p_{no}$ = thermal-equilibrium hole concentration in n side. From Eq. (2-15) $p_{no} = N_A$, and from Eq. (2-14) $p_{no} = n_i^2/N_D$, so that

$$V_o = V_T \ln \frac{N_A N_D}{n_i^2} \tag{2-63}$$

The same expression for V_o is obtained from an analysis corresponding to that given above and based upon equating the total electron current I_n to zero (Prob. 2-20). The p-n junction, both open-circuited and with an applied voltage, is studied in detail in Chaps. 3 and 19.

2-13 RECAPITULATION

The fundamental principles governing the electrical behavior of semi-conductors, discussed in this chapter, are summarized as follows:

1. Two types of mobile charge carriers (positive holes and negative electrons) are available. This bipolar nature of a semiconductor is to be contrasted with the unipolar property of a metal, which possesses only free electrons.
2. A semiconductor may be fabricated with donor (acceptor) impurities; so it contains mobile charges which are primarily electrons (holes).
3. The intrinsic concentration of carriers is a function of temperature. At room temperature, essentially all donors or acceptors are ionized.

4. Current is due to two distinct phenomena:

a. Carriers drift in an electric field (this conduction current is also available in a metal).

b. Carriers diffuse if a concentration gradient exists (a phenomenon which does not take place in a metal).

5. Carriers are continuously being generated (due to thermal creation of hole-electron pairs) and are simultaneously disappearing (due to recombination).

6. The fundamental law governing the flow of charge is called the *continuity equation*. It is formulated by considering that charge can neither be created nor destroyed if generation, recombination, drift, and diffusion are all taken into account.

7. If minority carriers (say, holes) are injected into a region containing majority carriers (say, an *n*-type bar), then usually the injected minority concentration is very small compared with the density of the majority carriers. For this low-level injection condition the minority current is predominantly due to diffusion; in other words, the *minority* drift current may be neglected.

8. The total majority-carrier flow is the sum of a drift and a diffusion current. The majority conduction current results from a small electric field internally created within the semiconductor because of the injected carriers.

9. The minority-carrier concentration injected into one end of a semiconductor bar decreases exponentially with distance into the specimen (as a result of diffusion and recombination).

10. Across an open-circuited *p-n* junction there exists a contact difference of potential.

These basic concepts are applied in the next chapter to the study of the *p-n* junction diode.

REFERENCES

1. Shockley, W.: "Electrons and Holes in Semiconductors," D. Van Nostrand Company, Inc., Princeton, N.J., reprinted February, 1963.

2. Adler, R. B., A. C. Smith, and R. L. Longini: "Introduction to Semiconductor Physics," vol.1, Semiconductor Electronics Education Committee, John Wiley & Sons, Inc., New York, 1964.

3. Morin, F. J., and J. P. Maita: Conductivity and Hall Effect in the Intrinsic Range of Germanium, *Phys. Rev.*, vol. 94, pp. 1525–1529, June, 1954.
Morin, F. J., and J. P. Maita: Electrical Properties of Silicon Containing Arsenic and Boron, *Phys. Rev.*, vol. 96, pp. 28–35, October, 1954.

4. Sze, S. M.: "Physics of Semiconductor Devices," Fig. 29, p. 59, John Wiley & Sons, Inc., New York, 1969.

5. Shockley, W., and W. T. Read, Jr.: Statistics of the Recombination of Holes and Electrons, *Phys. Rev.*, vol. 87, pp. 835–842, September, 1952.

Hall, R. N.: Electron-Hole Recombination in Germanium, *Phys. Rev.*, vol. 87, p. 387, July, 1952.

6. Collins, C. B., R. O. Carlson, and C. J. Gallagher: Properties of Gold-doped Silicon, *Phys. Rev.*, vol. 105, pp. 1168–1173, February, 1957.

Bemski, G.: Recombination Properties of Gold in Silicon, *Phys. Rev.*, vol. 111, pp. 1515–1518, September, 1958.

7. Gray, P. E., and C. L. Searle: "Electronic Principles: Physics, Models, and Circuits," John Wiley & Sons, Inc., New York 1969.

REVIEW QUESTIONS

2-1 Give the electron-gas description of a metal.

2-2 (*a*) Define *mobility*. (*b*) Give its dimensions.

2-3 (*a*) Define *conductivity*. (*b*) Give its dimensions.

2-4 Define a *hole* (in a semiconductor).

2-5 Indicate pictorially how a hole contributes to conduction.

2-6 (*a*) Define *intrinsic concentration* of holes. (*b*) What is the relationship between this density and the intrinsic concentration for electrons? (*c*) What do these equal at 0°K?

2-7 Show (in two dimensions) the crystal structure of silicon containing a donor impurity atom.

2-8 Repeat Rev. 2-7 for an acceptor impurity atom.

2-9 Define (*a*) *donor*, (*b*) *acceptor* impurities.

2-10 A semiconductor is doped with both donors and acceptors of concentrations N_D and N_A, respectively. Write the equation or equations from which to determine the electron and hole concentrations (n and p).

2-11 Define the *volt equivalent of temperature.*

2-12 Describe the *Hall effect.*

2-13 What properties of a semiconductor are determined from a Hall effect experiment?

2-14 Given an intrinsic semiconductor specimen, state two physical processes for increasing its conductivity. Explain briefly.

2-15 Is the temperature coefficient of resistance of a semiconductor positive or negative? Explain briefly.

2-16 Answer Rev. 2-15 for a metal.

2-17 (*a*) Sketch the spectral response curve for silicon. (*b*) Explain its shape qualitatively.

2-18 (*a*) Define *long-wavelength, threshold,* or *critical wavelength* for a semiconductor. (*b*) Explain why λ_c exists.

2-19 Define *mean lifetime* of a carrier.

2-20 Explain physically the meaning of the following statement: An electron and a hole recombine and disappear.

2-21 Radiation falls on a semiconductor specimen which is uniformly illuminated, and a steady-state is reached. At $t = 0$ the light is turned off. (*a*) Sketch the minority-carrier concentration as a function of time for $t \geq 0$. (*b*) Define all symbols in the equation describing your sketch.

2-22 (*a*) Define *diffusion constant* for holes. (*b*) Give its dimensions.

2-23 Repeat Rev. 2-22 for electrons.

2-24 (*a*) Write the equation for the net electron current in a semiconductor. What is the physical significance of each term? (*b*) How is this equation modified for a metal?

2-25 (*a*) The *equation of continuity* is a mathematical statement of what physical law? (*b*) The left-hand side of this equation for holes is dp/dt. The right-hand side contains several terms. State in words (no mathematics) what each of these terms represents physically.

2-26 Light falls upon the end of a long open-circuited semiconductor specimen. (*a*) Sketch the steady-state minority-carrier concentration as a function of distance. (*b*) Define all the symbols in the equation describing your sketch.

2-27 Light falls upon the end of a long open-circuited semiconductor bar. (*a*) For low-level injection is the minority current due predominantly to drift, diffusion, or both? (*b*) Is the majority current due predominantly to drift, diffusion, or both?

2-28 (*a*) Define a *graded semiconductor*. (*b*) Explain why an electric field must exist in a graded semiconductor.

2-29 Consider a step-graded junction under open-circuited conditions. Upon what four parameters does the contact difference of potential depend?

2-30 State the *mass-action law* as an equation and in words.

2-31 Explain why a contact difference of potential must develop across an open-circuited *p-n* junction.

3 / JUNCTION-DIODE CHARACTERISTICS

In this chapter we demonstrate that if a junction is formed between a sample of p-type and one of n-type semiconductor, this combination possesses the properties of a rectifier. The volt-ampere characteristics of such a two-terminal device (called a *junction diode*) is studied. The capacitance across the junction is calculated.

Although the transistor is a triode (three-terminal) semiconductor, it may be considered as one diode biased by the current from a second diode. Hence most of the theory developed here is utilized in Chap. 5 in connection with the study of the transistor.

3-1 THE OPEN-CIRCUITED p-n JUNCTION

If donor impurities are introduced into one side and acceptors into the other side of a single crystal of a semiconductor, a p-n junction is formed, as in Fig. 2-17*b*. Such a system is illustrated in more schematic detail in Fig. 3-1*a*. The donor ion is represented by a plus sign because, after this impurity atom "donates" an electron, it becomes a positive ion. The acceptor ion is indicated by a minus sign because, after this atom "accepts" an electron, it becomes a negative ion. Initially, there are nominally only p-type carriers to the left of the junction and only n-type carriers to the right.

Space-charge Region Because there is a density gradient across the junction, holes will initially diffuse to the right across the junction, and electrons to the left. We see that the positive holes which neutralized the acceptor ions near the junction in the p-type silicon have disappeared as a result of combination with electrons which have diffused across the junction. Similarly, the neutralizing electrons in

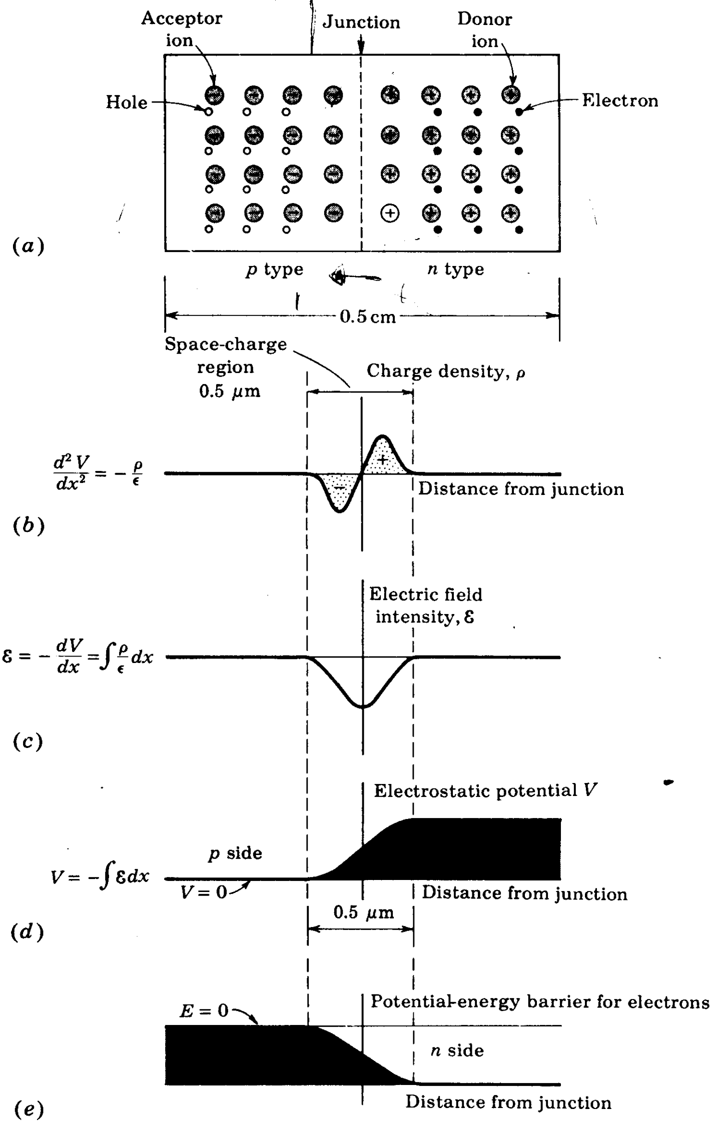

Acceptor ion
Junction
Donor ion
Hole
Electron
(a)
p type
n type
0.5 cm
Space-charge region 0.5 μm
Charge density, ρ
$\frac{d^2V}{dx^2} = -\frac{\rho}{\epsilon}$
(b)
Distance from junction
Electric field intensity, ℰ
$\mathcal{E} = -\frac{dV}{dx} = \int \frac{\rho}{\epsilon} dx$
(c)
Electrostatic potential V
p side
$V = -\int \mathcal{E} dx$
V = 0
Distance from junction
0.5 μm
(d)
E = 0
Potential-energy barrier for electrons
n side
Distance from junction
(e)

Fig. 3-1 A schematic diagram of a p-n junction, including the charge density, electric field intensity, and potential-energy barriers at the junction. Since potential energy = potential × charge, the curve in (*d*) is proportional to the potential energy for a hole (a positive charge) and the curve in (e) is proportional to the negative of that in (*d*) (an electron is a negative charge). (Not drawn to scale.)

the *n*-type silicon have combined with holes which have crossed the junction from the *p* material. The unneutralized ions in the neighborhood of the junction are referred to as *uncovered charges*. The general shape of the charge density ρ (Fig. 3-1*b*) depends upon how the diode is doped (a step-graded junction is considered in detail in Sec. 3-7). Since the region of the junction is depleted of mobile charges, it is called the *depletion region*, the *space-charge region*, or the *transition region*. The thickness of this region is of the order of the wavelength of visible light (0.5 micron = 0.5 μm). Within this very narrow space-charge layer there are no mobile carriers. To the left of this region the carrier concentration is $p \approx N_A$, and to its right it is $n \approx N_D$.

Electric Field Intensity The space-charge density ρ is zero at the junction. It is positive to the right and negative to the left of the junction. This distribution constitutes an electrical dipole layer, giving rise to electric lines of flux from right to left, corresponding to negative field intensity $\mathcal{E}$ as depicted in Fig. 3-1*c*. Equilibrium is established when the field is strong enough to restrain the process of diffusion. Stated alternatively, under steady-state conditions the drift hole (electron) current must be equal and opposite to the diffusion hole (electron) current so that the net hole (electron) current is reduced to zero—as it must be for an open-circuited device. In other words, there is no steady-state movement of charge across the junction.

The field intensity curve is proportional to the integral of the charge density curve. This statement follows from Poisson's equation

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\epsilon} \qquad (3\text{-}1)$$

where ϵ is the permittivity. If ϵ_r is the (relative) dielectric constant and ϵ_o is the permittivity of free space (Appendix A), then $\epsilon = \epsilon_r\epsilon_o$. Integrating Eq. (3-1) and remembering that $\mathcal{E} = -dV/dx$ gives

$$\mathcal{E} = \int_{x_o}^{x} \frac{\rho}{\epsilon}\,dx \qquad (3\text{-}2)$$

where $\mathcal{E} = 0$ at $x = x_o$. Therefore the curve plotted in Fig. 3-1*c* is the integral of the function drawn in Fig. 3-1*b* (divided by ϵ).

Potential The electrostatic-potential variation in the depletion region is shown in Fig. 3-1*d*, and is the negative integral of the function $\mathcal{E}$ of Fig. 3-1*c*. This variation constitutes a potential-energy barrier (Sec. 1-2) against the further diffusion of holes across the barrier. The form of the potential-energy barrier against the flow of electrons from the *n* side across the junction is shown in Fig. 3-1*e*. It is similar to that shown in Fig. 3-1*d*, except that it is inverted, since the charge on an electron is negative. Note the existence, across the depletion layer, of the *contact potential* V_o, discussed in Sec. 2-12.

Summary Under open-circuited conditions the net hole current must be zero. If this statement were not true, the hole density at one end of the semiconductor would continue to increase indefinitely with time, a situation which is obviously physically impossible. Since the concentration of holes in the p side is much greater than that in the n side, a very large hole diffusion current tends to flow across the junction from the p to the n material. Hence an electric field must build up across the junction in such a direction that a hole drift current will tend to flow across the junction from the n to the p side in order to counterbalance the diffusion current. This equilibrium condition of zero resultant hole current allows us to calculate the height of the potential barrier V_o [Eq. (2-63)] in terms of the donor and acceptor concentrations. The numerical value for V_o is of the order of magnitude of a few tenths of a volt.

3-2 THE p-n JUNCTION AS A RECTIFIER[1]

The essential electrical characteristic of a p-n junction is that it constitutes a rectifier which permits the easy flow of charge in one direction but restrains the flow in the opposite direction. We consider now, qualitatively, how this diode rectifier action comes about.

Reverse Bias In Fig. 3-2, a battery is shown connected across the terminals of a p-n junction. The negative terminal of the battery is connected to the p side of the junction, and the positive terminal to the n side. The polarity of connection is such as to cause both the holes in the p type and the electrons in the n type to move away from the junction. Consequently, the region of negative-charge density is spread to the left of the junction (Fig. 3-1b), and the positive-charge-density region is spread to the right. However, this process cannot continue indefinitely, because in order to have a steady flow of holes to the left, these holes must be supplied across the junction from the n-type silicon. And there are very few holes in the n-type side. Hence, nominally, zero current results. Actually, a small current does flow because a small number of hole-electron pairs are generated throughout the crystal as a result of thermal energy. The holes so formed in the n-type silicon will wander over to the junction. A similar remark applies to the electrons

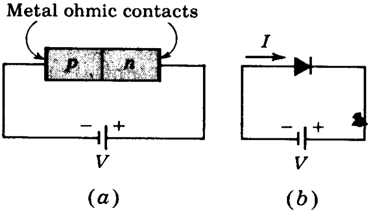
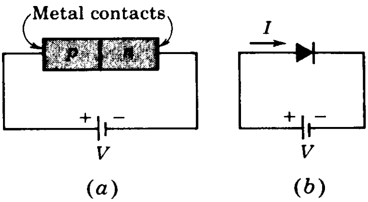


Fig. 3-2 (a) A p-n junction biased in the reverse direction. (b) The rectifier symbol is used for the p-n diode.

Fig. 3-3 (*a*) A p-n junction biased in the forward direction. (*b*) The rectifier symbol is used for the p-n diode.



thermally generated in the p-type silicon. This small current is the diode *reverse saturation current*, and its magnitude is designated by I_o. This reverse current will increase with increasing temperature [Eq. (3-11)], and hence the back resistance of a crystal diode decreases with increasing temperature. From the argument presented here, I_o should be independent of the magnitude of the reverse bias.

The mechanism of conduction in the reverse direction may be described alternatively in the following way: When no voltage is applied to the p-n diode, the potential barrier across the junction is as shown in Fig. 3-1*d*. When a voltage V is applied to the diode in the direction shown in Fig. 3-2, the height of the potential-energy barrier is increased by the amount qV. This increase in the barrier height serves to reduce the flow of majority carriers (i.e., holes in p type and electrons in n type). However, the minority carriers (i.e., electrons in p type and holes in n type), since they fall down the potential-energy hill, are uninfluenced by the increased height of the barrier. The applied voltage in the direction indicated in Fig. 3-2 is called the *reverse*, or *blocking*, *bias*.

Forward Bias An external voltage applied with the polarity shown in Fig. 3-3 (opposite to that indicated in Fig. 3-2) is called a *forward* bias. An ideal p-n diode has zero ohmic voltage drop across the body of the crystal. For such a diode the height of the potential barrier at the junction will be lowered by the applied forward voltage V. The equilibrium initially established between the forces tending to produce diffusion of majority carriers and the restraining influence of the potential-energy barrier at the junction will be disturbed. Hence, for a forward bias, the holes cross the junction from the p-type into the n-type region, where they constitute an injected minority current. Similarly, the electrons cross the junction in the reverse direction and become a minority current injected into the p side. Holes traveling from left to right constitute a current in the *same* direction as electrons moving from right to left. Hence the resultant current crossing the junction is the *sum* of the hole and electron minority currents. A detailed discussion of the several current components within the diode is given in the next section.

Ohmic Contacts In Fig. 3-2 (3-3) we show an external reverse (forward) bias applied to a p-n diode. We have assumed that the external bias voltage appears directly across the junction and has the effect of raising (lowering) the electrostatic potential across the junction. To justify this assumption we must specify how electric contact is made to the semiconductor from the external bias circuit. In Figs. 3-2 and 3-3 we indicate metal contacts with which the homogeneous p-type and n-type materials are provided. We thus see that we have introduced two metal-semiconductor junctions, one at each end of the diode. We naturally expect a contact potential to develop across these additional junctions. However, we shall assume that the metal-semiconductor contacts shown in Figs. 3-2 and 3-3 have been manufactured in such a way that they are nonrectifying. In other words, the contact potential across these junctions is constant, independent of the direction and magnitude of the current. A contact of this type is referred to as an *ohmic contact.*

We are now in a position to justify our assumption that the entire applied voltage appears as a *change* in the height of the potential barrier. Inasmuch as the voltage across the metal-semiconductor ohmic contacts remains constant and the voltage drop across the bulk of the crystal is neglected, approximately the entire applied voltage will indeed appear as a change in the height of the potential barrier at the p-n junction.

The Short-circuited and Open-circuited p-n Junction If the voltage V in Fig. 3-2 or 3-3 were set equal to zero, the p-n junction would be short-circuited. Under these conditions, as we show below, no current can flow ($I = 0$) and the electrostatic potential V_o remains unchanged and equal to the value under open-circuit conditions. If there were a current ($I \neq 0$), the metal would become heated. Since there is no external source of energy available, the energy required to heat the metal wire would have to be supplied by the p-n bar. The semiconductor bar, therefore, would have to cool off. Clearly, under thermal equilibrium the simultaneous heating of the metal and cooling of the bar is impossible, and we conclude that $I = 0$. Since under short-circuit conditions the sum of the voltages around the closed loop must be zero, the junction potential V_o must be exactly compensated by the metal-to-semiconductor contact potentials at the ohmic contacts. Since the current is zero, the wire can be cut without changing the situation, and the voltage drop across the cut must remain zero. If in an attempt to measure V_o we connected a voltmeter across the cut, the voltmeter would read zero voltage. In other words, it is not possible to measure contact difference of potential directly with a voltmeter.

Large Forward Voltages Suppose that the forward voltage V in Fig. 3-3 is increased until V approaches V_o. If V were equal to V_o, the barrier would disappear and the current could be arbitrarily large, exceeding the rating of the diode. As a practical matter we can never reduce the barrier to zero because, as the current increases without limit, the bulk resistance of the crystal, as well as the resistance of the ohmic contacts, will limit the current. Therefore it is no longer possible to assume that all the voltage V appears as a change across the p-n junction. We conclude that, as the forward voltage V becomes comparable with V_o, the current through a real p-n diode will be governed by the ohmic-contact resistances and the crystal bulk resistance. Thus the volt-ampere characteristic becomes approximately a straight line.

3-3 THE CURRENT COMPONENTS IN a p-n DIODE

In the preceding section it was indicated that when a forward bias is applied to a diode, holes are injected into the n side and electrons into the p side. In Sec. 2-11 it was emphasized that under low-level injection conditions such minority currents are due almost entirely to diffusion, so that minority drift currents may be neglected. From Eq. (2-48) the hole diffusion current in the n-type material† I_{pn} decreases exponentially with distance x into the n-type region and falls to $1/\epsilon$th its peak value in a diffusion length L_p. This current is plotted in Fig. 3-4, as is also the corresponding electron diffusion current I_{np} in the p-type side. The doping on the two sides of the junction need not be identical, and it is assumed in this plot that the acceptor concentration is much greater than the donor density, so that the hole current greatly exceeds

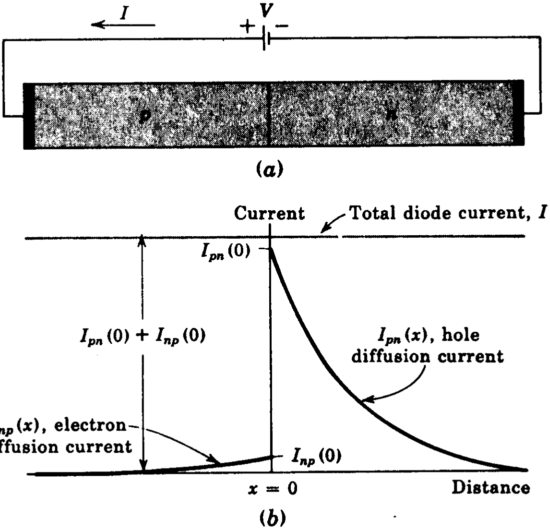


Fig. 3-4 The hole- and electron-current diffusion components vs. distance in a p-n junction diode. The p side is much more heavily doped than the n section. The space-charge region at the junction is assumed to be negligibly small.

---

† Since we must consider both hole and electron currents in each side of the junction, we add the second subscript n to I_p in order to indicate that the hole current in the n-type region is under consideration. In general, if the letters p and n both appear in a symbol, the first letter refers to the type of carrier and the second to the type of material.

the electron current. Also, in Fig. 3-4, the space-charge region has been assumed to be negligibly small, a restriction to be removed soon.

From Eq. (2-48) (with the subscript n added to I_p and to p) the minority (hole) diffusion current at the junction ($x = 0$) is given by

$$I_{pn}(0) = \frac{AqD_p}{L_p}[p_n(0) - p_{no}] \tag{3-3}$$

The Law of the Junction In the preceding section it was pointed out that a forward bias V lowers the barrier height and allows more carriers to cross the junction. Hence $p_n(0)$ must be a function of V. From the Boltzmann relationship, Eq. (2-59), it seems reasonable that $p_n(0)$ should depend exponentially, upon V. Indeed, in Sec. 19-10, it is found that

$$p_n(0) = p_{no}\epsilon^{V/V_T} \tag{3-4}$$

This relationship, which gives the hole concentration at the edge of the n region (at $x = 0$, just outside of the transition region) in terms of the thermal-equilibrium minority-carrier concentration p_{no} (far away from the junction) and the applied potential V, is called the *law of the junction.* A similar equation with p and n interchanged gives the electron concentration at the edge of the p region in terms of V.

The Total Diode Current Substituting Eq. (3-4) into Eq. (3-3) yields

$$I_{pn}(0) = \frac{AqD_p p_{no}}{L_p}(\epsilon^{V/V_T} - 1) \tag{3-5}$$

The expression for the electron current $I_{np}(0)$ crossing the junction into the p side is obtained from Eq. (3-5) by interchanging p and n.

Electrons crossing the junction at $x = 0$ from right to left constitute a current in the same direction as holes crossing the junction from left to right. Hence the total diode current I at $x = 0$ is

$$I = I_{pn}(0) + I_{np}(0) \tag{3-6}$$

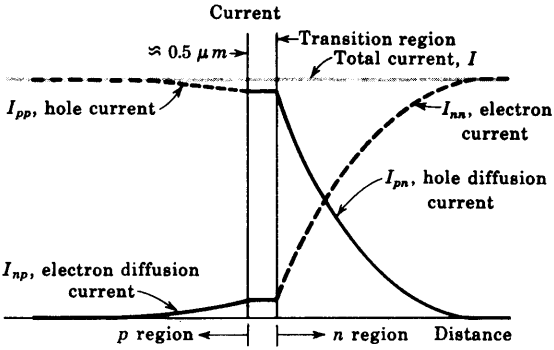
Since the current is the same throughout a series circuit, I is independent of x and is indicated as a horizontal line in Fig. 3-4*b*. The expression for the diode current is

$$I = I_o(\epsilon^{V/V_T} - 1) \tag{3-7}$$

where I_o is given in Prob. 3-6 in terms of the physical parameters of the diode.

The Reverse Saturation Current In the foregoing discussion a positive value of V indicates a forward bias. The derivation of Eq. (3-7) is equally valid if V is negative, signifying an applied reverse-bias voltage. For a reverse bias whose magnitude is large compared with V_T ($\approx$26 mV at room temperature), $I \rightarrow -I_o$. Hence I_o is called the *reverse saturation current.*

Fig. 3-5 The minority (solid) and the majority (dashed) currents vs. distance in a *p-n* diode. It is assumed that no recombination takes place in the very narrow depletion region.



Since the thermal-equilibrium concentrations p_{no} and n_{po} depend upon temperature T, then I_o is a function of T. This temperature dependence is derived in Sec. 19-10.

The Majority-carrier Current Components In the n-type region of Fig. 3-4*b* the total current I is constant and the minority (hole) current I_{pn} varies with x. Clearly, there must exist a majority (electron) current I_{nn} which is a function of x because the diode current I at any position is the sum of the hole and electron currents at this distance. This majority current

$$I_{nn}(x) = I - I_{pn}(x) \tag{3-8}$$

is plotted as the dashed curve in the n region of Fig. 3-5, where the narrow depletion region is also indicated. The majority hole current I_{pp} is similarly shown dashed in the p region of this figure. As discussed in Sec. 2-11, these majority currents are each composed of two current components; one is a drift current and the second is a diffusion current. Recall from Sec. 2-11 that the diffusion electron current is $-(D_n/D_p)I_{pn}$ in the n side.

Note that deep into the p side the current is a drift (conduction) current I_{pp} of holes sustained by the small electric field in the semiconductor (Prob. 3-4). As the holes approach the junction, some of them recombine with the electrons, which are injected into the p side from the n side. The current I_{pp} thus decreases toward the junction (at just the proper gradient to maintain the total current constant, independent of distance). What remains of I_{pp} at the junction enters the n side and becomes the hole diffusion current I_{pn}. Similar remarks can be made with respect to current I_{nn}.

We emphasize that the current in a p-n diode is bipolar in character since it is made up of both positive and negative carriers of electricity. The total current is constant throughout the device, but the proportion due to holes and that due to electrons varies with distance, as indicated in Fig. 3-5.

The Transition Region Since this depletion layer contains very few mobile charges, it has been assumed (Fig. 3-5) that carrier generation and recombination may be neglected within the bulk and on the surface of this region. Such an assumption is valid for a germanium diode, but not for a silicon device. For the latter it is found[2] that Eq. (3-5) must be modified by multiplying V_T by a factor η, where $\eta \approx 2$ for small (rated) currents and $\eta \approx 1$ for large currents.

3-4 THE VOLT-AMPERE CHARACTERISTIC

The discussion of the preceding section indicates that, for a *p-n* junction, the current I is related to the voltage V by the equation

$$I = I_o(\epsilon^{V/\eta V_T} - 1) \tag{3-9}$$

A positive value of I means that current flows from the p to the n side. The diode is forward-biased if V is positive, indicating that the p side of the junction is positive with respect to the n side. The symbol η is unity for germanium and is approximately 2 for silicon at rated current.

The symbol V_T stands for the volt equivalent of temperature, and is given by Eq. (2-38), repeated here for convenience:

$$V_T \equiv \frac{T}{11{,}600} \tag{3-10}$$

At room temperature ($T = 300°\text{K}$), $V_T = 0.026\ \text{V} = 26\ \text{mV}$.

The form of the volt-ampere characteristic described by Eq. (3-9) is shown in Fig. 3-6*a*. When the voltage V is positive and several times V_T,

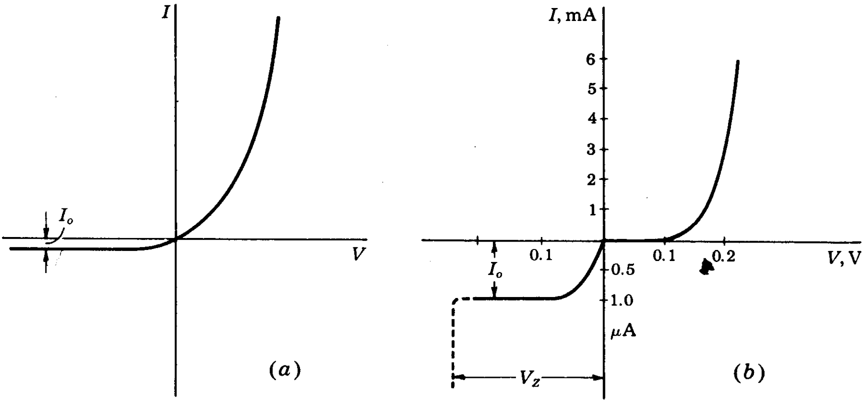


Fig. 3-6 (a) The volt-ampere characteristic of an ideal *p-n* diode. (b) The volt-ampere characteristic for a germanium diode redrawn to show the order of magnitude of currents. Note the expanded scale for reverse currents. The dashed portion indicates breakdown at V_Z.

the unity in the parentheses of Eq. (3-9) may be neglected. Accordingly, except for a small range in the neighborhood of the origin, the current increases exponentially with voltage. When the diode is reverse-biased and $|V|$ is several times V_T, $I \approx -I_o$. The reverse current is therefore constant, independent of the applied reverse bias. Consequently, I_o is referred to as the *reverse saturation current.*

For the sake of clarity, the current I_o in Fig. 3-6 has been greatly exaggerated in magnitude. Ordinarily, the range of forward currents over which a diode is operated is many orders of magnitude larger than the reverse saturation current. To display forward and reverse characteristics conveniently, it is necessary, as in Fig. 3-6*b*, to use two different current scales. The volt-ampere characteristic shown in that figure has a forward-current scale in milliamperes and a reverse scale in microamperes.

The dashed portion of the curve of Fig. 3-6*b* indicates that, at a reverse-biasing voltage V_Z, the diode characteristic exhibits an abrupt and marked departure from Eq. (3-9). At this critical voltage a large reverse current flows, and the diode is said to be in the *breakdown* region, discussed in Sec. 3-11.

The Cutin Voltage V_γ Both silicon and germanium diodes are commercially available. A number of differences between these two types are relevant in circuit design. The difference in volt-ampere characteristics is brought out in Fig. 3-7. Here are plotted the forward characteristics at room temperature of a general-purpose germanium switching diode and a general-purpose silicon diode, the 1N270 and 1N3605, respectively. The diodes have comparable current ratings. A noteworthy feature in Fig. 3-7 is that there exists a *cutin*, *offset*, *break-point*, or *threshold*, voltage V_γ below which the current is very small (say, less than 1 percent of maximum rated value). Beyond

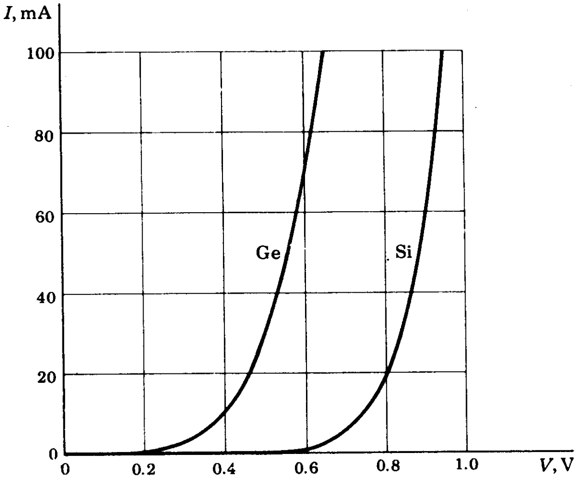


Fig. 3-7 The forward volt-ampere characteristics of a germanium (1N270) and a silicon (1N3605) diode at 25°C.

V_γ the current rises very rapidly. From Fig. 3-7 we see that V_γ is approximately 0.2 V for germanium and 0.6 V for silicon.

Note that the break in the silicon-diode characteristic is offset about 0.4 V with respect to the break in the germanium-diode characteristic. The reason for this difference is to be found, in part, in the fact that the reverse saturation current in a germanium diode is normally larger by a factor of about 1,000 than the reverse saturation current in a silicon diode of comparable ratings. I_o is in the range of microamperes for a germanium diode and nanoamperes for a silicon diode at room temperature.

Since $\eta = 2$ for small currents in silicon, the current increases as $\epsilon^{V/2V_T}$ for the first several tenths of a volt and increases as ϵ^{V/V_T} only at higher voltages. This initial smaller dependence of the current on voltage accounts for the further delay in the rise of the silicon characteristic.

Logarithmic Characteristic It is instructive to examine the family of curves for the silicon diodes shown in Fig. 3-8. A family for a germanium diode of comparable current rating is quite similar, with the exception that corresponding currents are attained at lower voltage.

From Eq. (3-9), assuming that V is several times V_T, so that we may drop the unity, we have $\log I = \log I_o + 0.434V/\eta V_T$. We therefore expect in Fig. 3-8, where $\log I$ is plotted against V, that the plots will be straight lines. We do indeed find that at low currents the plots are linear and correspond to $n \approx 2$. At large currents an increment of voltage does not yield as large an increase of current as at low currents. The reason for this behavior is to be found in the ohmic resistance of the diode. At low currents the ohmic drop is negligible and the externally impressed voltage simply decreases the potential barrier at the p-n junction. At high currents the externally impressed voltage is called upon principally to establish an electric field to

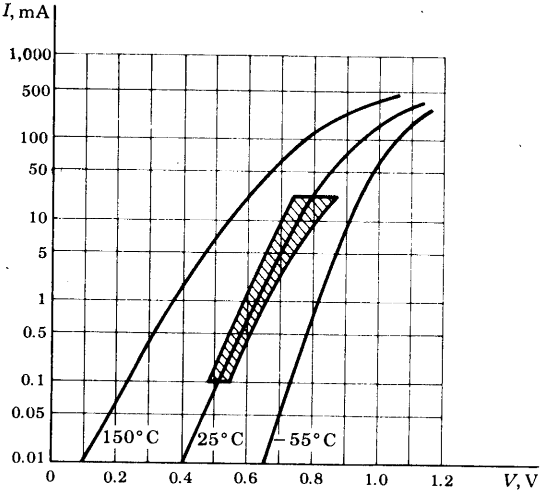


Fig. 3-8 Volt-ampere characteristics at three different temperatures for a silicon diode (planar epitaxial passivated types 1N3605, 1N3606, 1N3608, and 1N3609). The shaded area indicates 25°C limits of controlled conductance. Note that the vertical scale is logarithmic and encompasses a current range of 50,000. (Courtesy of General Electric Company.)

overcome the ohmic resistance of the semiconductor material. Therefore, at high currents, the diode behaves more like a resistor than a diode, and the current increases linearly rather than exponentially with applied voltage.

Reverse Saturation Current Many commercially available diodes exhibit an essentially constant value of I_o for negative values of V, as indicated in Fig. 3-6. On the other hand, some diodes show a very pronounced increase in reverse current with increasing reverse voltage. This variation in I_o results from leakage across the surface of the diode, and also from the additional fact that new charge carriers may be generated by collision in the transition region at the junction.

3-5 THE TEMPERATURE DEPENDENCE OF THE V/I CHARACTERISTIC

The volt-ampere relationship (3-9) contains the temperature implicitly in the two symbols V_T and I_o. In Sec. 19-11 it is shown that the theoretical variation of I_o with T is 8 percent/°C for silicon and 11 percent/°C for germanium. The performance of commercial diodes is only approximately consistent with these results. The reason for the discrepancy is that, in a physical diode, there is a component of the reverse saturation current due to leakage over the surface that is not taken into account in Sec. 19-11. Since this leakage component is independent of temperature, we may expect to find a smaller rate of change of I_o with temperature than that predicted above. From experimental data we observe that the reverse saturation current increases approximately 7 percent/°C for both silicon and germanium. Since $(1.07)^{10} \approx 2.0$, we conclude that *the reverse saturation current approximately doubles for every* 10°C *rise in temperature*. If $I_o = I_{o1}$ at $T = T_1$, then at a temperature T, I_o is given by

$$I_o(T) = I_{o1} \times 2^{(T-T_1)/10} \tag{3-11}$$

If the temperature is increased at a fixed voltage, the current increases. However, if we now reduce V, then I may be brought back to its previous value. In Sec. 19-11 it is found that for either silicon or germanium (*at room temperature*)

$$\frac{dV}{dT} \approx -2.5 \text{ mV/°C} \tag{3-12}$$

in order to maintain a constant value of I. It should also be noted that $|dV/dT|$ decreases with increasing T.

3-6 DIODE RESISTANCE

The static resistance R of a diode is defined as the ratio V/I of the voltage to the current. At any point on the volt-ampere characteristic of the diode

(Fig. 3-7), the resistance R is equal to the reciprocal of the slope of a line joining the operating point to the origin. The static resistance varies widely with V and I and is not a useful parameter. The rectification property of a diode is indicated on the manufacturer's specification sheet by giving the maximum forward voltage V_F required to attain a given forward current I_F and also the maximum reverse current I_R at a given reverse voltage V_R. Typical values for a silicon planar epitaxial diode are $V_F = 0.8$ V at $I_F = 10$ mA (corresponding to $R_F = 80\ \Omega$) and $I_R = 0.1\ \mu$A at $V_R = 50$ V (corresponding to $R_R = 500$ M).

For small-signal operation the *dynamic*, or *incremental, resistance* r is an important parameter, and is defined as the reciprocal of the slope of the volt-ampere characteristic, $r \equiv dV/dI$. The dynamic resistance is not a constant, but depends upon the operating voltage. For example, for a semiconductor diode, we find from Eq. (3-9) that the dynamic conductance $g \equiv 1/r$ is

$$g \equiv \frac{dI}{dV} = \frac{I_o \epsilon^{V/\eta V_T}}{\eta V_T} = \frac{I + I_o}{\eta V_T} \tag{3-13}$$

For a reverse bias greater than a few tenths of a volt (so that $|V/\eta V_T| \gg 1$), g is extremely small and r is very large. On the other hand, for a forward bias greater than a few tenths of a volt, $I \gg I_o$, and r is given approximately by

$$r \approx \frac{\eta V_T}{I} \tag{3-14}$$

The dynamic resistance varies inversely with current; at room temperature and for $\eta = 1$, $r = 26/I$, where I is in milliamperes and r in ohms. For a forward current of 26 mA, the dynamic resistance is 1 Ω. The ohmic body resistance of the semiconductor may be of the same order of magnitude or even much higher than this value. Although r varies with current, in a small-signal model, it is reasonable to use the parameter r as a constant.

A Piecewise Linear Diode Characteristic A large-signal approximation which often leads to a sufficiently accurate engineering solution is the *piecewise linear* representation. For example, the piecewise linear approximation for a semiconductor diode characteristic is indicated in Fig. 3-9. The break point is not at the origin, and hence V_γ is also called the *offset*, or *threshold, voltage*. The diode behaves like an open circuit if $V < V_\gamma$, and has a constant incremental resistance $r = dV/dI$ if $V > V_\gamma$. Note that the resistance r (also designated as R_f and called the *forward resistance*) takes on added physical significance even for this large-signal model, whereas the static resistance $R_F = V/I$ is not constant and is not useful.

The numerical values V_γ and R_f to be used depend upon the type of diode and the contemplated voltage and current swings. For example, from Fig. 3-7 we find that, for a current swing from cutoff to 10 mA with a germanium diode, reasonable values are $V_\gamma = 0.2$ V and $R_f = 20\ \Omega$, and for a silicon diode, $V_\gamma = 0.6$ V and $R_f = 15\ \Omega$. On the other hand, a better approximation

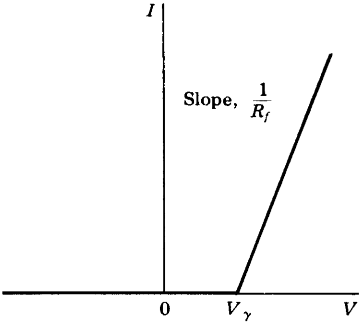


Fig. 3-9 The piecewise linear characterization of a semiconductor diode.

for current swings up to 50 mA leads to the following values; germanium, $V_\gamma = 0.3$ V, $R_f = 6\ \Omega$; silicon, $V_\gamma = 0.65$ V, $R_f = 5.5\ \Omega$. For an avalanche diode, discussed in Sec. 3-11, $V_\gamma = V_Z$ and R_f is the dynamic resistance in the breakdown region.

3-7 SPACE-CHARGE, OR TRANSITION, CAPACITANCE C_T

As mentioned in Sec. 3-1, a reverse bias causes majority carriers to move away from the junction, thereby uncovering more immobile charges. Hence the thickness of the space-charge layer at the junction increases with reverse voltage. This increase in uncovered charge with applied voltage may be considered a capacitive effect. We may define an incremental capacitance C_T by

$$C_T = \left| \frac{dQ}{dV} \right| \tag{3-15}$$

where dQ is the increase in charge caused by a change dV in voltage. It follows from this definition that a change in voltage dV in a time dt will result in a current $i = dQ/dt$, given by

$$i = C_T \frac{dV}{dt} \tag{3-16}$$

Therefore a knowledge of C_T is important in considering a diode (or a transistor) as a circuit element. The quantity C_T is referred to as the *transition-region, space-charge, barrier*, or *depletion-region, capacitance.* We now consider C_T quantitatively. As it turns out, this capacitance is not a constant, but depends upon the magnitude of the reverse voltage. It is for this reason that C_T is defined by Eq. (3-16) rather than as the ratio Q/V.

A Step-graded Junction Consider a junction in which there is an abrupt change from acceptor ions on one side to donor ions on the other side. Such a junction is formed experimentally, for example, by placing indium, which is trivalent, against n-type germanium and heating the combination to a high

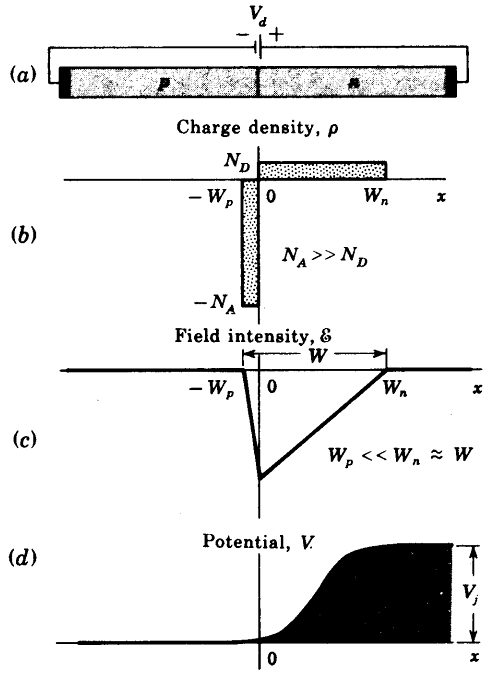


Fig. 3-10 (a) A reverse-biased *p-n* step-graded junction. (b) The charge density. (c) The field intensity. (d) The potential variation with distance *x*.

temperature for a short time. Some of the indium dissolves into the germanium to change the germanium from *n* to *p* type at the junction. Such a step-graded junction is called an *alloy*, or *fusion*, *junction*. A step-graded junction is also formed between emitter and base of an integrated transistor (Fig. 7-11). It is not necessary that the concentration N_A of acceptor ions equal the concentration N_D of donor impurities. As a matter of fact, it is often advantageous to have an unsymmetrical junction. Figure 3-10 shows the charge density as a function of distance from an alloy junction in which the acceptor impurity density is assumed to be much larger than the donor concentration. Since the net charge must be zero, then

$$N_A W_p = N_D W_n \tag{3-17}$$

If $N_A \gg N_D$, then $W_p \ll W_n \approx W$. The relationship between potential and charge density is given by Eq. (3-1):

$$\frac{d^2V}{dx^2} = \frac{-qN_D}{\epsilon} \tag{3-18}$$

The electric lines of flux start on the positive donor ions and terminate on the negative acceptor ions. Hence there are no flux lines to the right of the boundary $x = W_n$ in Fig. 3-10, and $\mathcal{E} = -dV/dx = 0$ at $x = W_n \approx W$.

Integrating Eq. (3-18) subject to this boundary condition yields

$$\frac{dV}{dx} = \frac{-qN_D}{\epsilon}(x - W) = -\mathcal{E} \tag{3-19}$$

Neglecting the small potential drop across W_p, we may arbitrarily choose $V = 0$ at $x = 0$. Integrating Eq. (3-19) subject to this condition gives

$$V = \frac{-qN_D}{2\epsilon}(x^2 - 2Wx) \tag{3-20}$$

The linear variation in field intensity and the quadratic dependence of potential upon distance are plotted in Fig. 3-10*c* and *d*. These graphs should be compared with the corresponding curves of Fig. 3-1.

At $x = W$, $V = V_j$ = junction, or barrier, potential. Thus

$$V_j = \frac{qN_D W^2}{2\epsilon} \tag{3-21}$$

In this section we have used the symbol V to represent the potential at any distance x from the junction. Hence, let us introduce V_d as the externally applied diode voltage. Since the barrier potential represents a reverse voltage, it is lowered by an applied forward voltage. Thus

$$V_j = V_o - V_d$$

where V_d is a negative number for an applied *reverse* bias and V_o is the contact potential (Fig. 3-1*d*). This equation confirms our qualitative conclusion that the thickness of the depletion layer increases with applied reverse voltage. We now see that W varies as $V_j^{\frac{1}{2}} = (V_o - V_d)^{\frac{1}{2}}$.

If A is the area of the junction, the charge in the distance W is

$$Q = qN_D WA$$

The transition capacitance C_T, given by Eq. (3-15), is

$$C_T = \left|\frac{dQ}{dV_d}\right| = qN_D A \left|\frac{dW}{dV_j}\right| \tag{3-22}$$

From Eq. (3-21), $|dW/dV_j| = \epsilon/qN_D W$, and hence

$$C_T = \frac{\epsilon A}{W} \tag{3-23}$$

It is interesting to note that this formula is exactly the expression which is obtained for a parallel-plate capacitor of area A (square meters) and plate separation W (meters) containing a material of permittivity ϵ. If the concentration N_A is not neglected, the above results are modified only slightly. In Eq. (3-21) W represents the total space-charge width, and $1/N_D$ is replaced by $1/N_A + 1/N_D$. Equation (3-23) remains valid (Prob. 3-18).

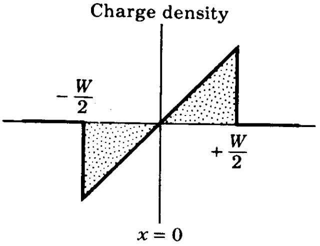


Fig. 3-11 The charge-density variation vs. distance at a linearly graded p-n junction.

A Linearly Graded Junction A second form of junction is obtained by drawing a single crystal from a melt of germanium whose type is changed during the drawing process by adding first p-type and then n-type impurities. A linearly graded junction is also formed between the collector and base of an integrated transistor (Fig. 7-12). For such a junction the charge density varies gradually (almost linearly), as indicated in Fig. 3-11. If an analysis similar to that given above is carried out for such a junction, Eq. (3-23) is found to be valid where W equals the total width of the space-charge layer. However, it now turns out that W varies as $V_j^{\frac{1}{3}}$ instead of $V_j^{\frac{1}{2}}$.

Varactor Diodes We observe from the above equations that the barrier capacitance is not a constant but varies with applied voltage. The larger the reverse voltage, the larger is the space-charge width W, and hence the smaller the capacitance C_T. The variation is illustrated for two typical diodes in Fig. 3-12. Similarly, for an increase in forward bias (V_d positive), W decreases and C_T increases.

The voltage-variable capacitance of a p-n junction biased in the reverse direction is useful in a number of circuits. One of these applications is voltage tuning of an LC resonant circuit. Other applications include self-balancing bridge circuits and special types of amplifiers, called *parametric amplifiers*.

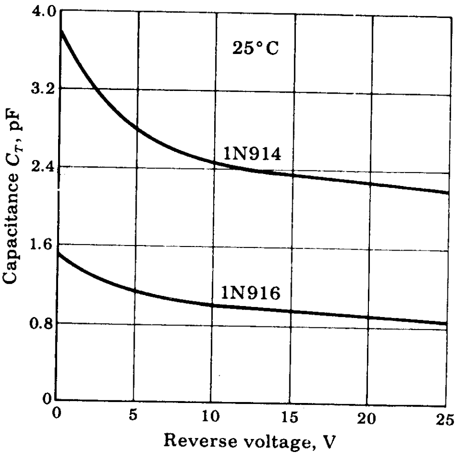


Fig. 3-12 Typical barrier-capacitance variation, with reverse voltage, of silicon diodes 1N914 and 1N916. (Courtesy of Fairchild Semiconductor Corporation.)

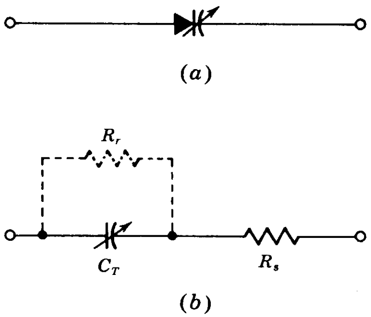


Fig. 3-13 A varactor diode under reverse bias. (a) Circuit symbol; (b) circuit model.

Diodes made especially for the above applications which are based on the voltage-variable capacitance are called *varactors, varicaps*, or *voltacaps*. A circuit model for a varactor diode under reverse bias is shown in Fig. 3-13. The resistance R_s represents the body (ohmic) series resistance of the diode. Typical values of C_T and R_s are 20 pF and 8.5 Ω, respectively, at a reverse bias of 4 V. The reverse diode resistance R_r shunting C_T is large (>1 M), and hence is usually neglected.

In circuits intended for use with fast waveforms or at high frequencies, it is required that the transition capacitance be as small as possible, for the following reason: a diode is driven to the reverse-biased condition when it is desired to prevent the transmission of a signal. However, if the barrier capacitance C_T is large enough, the current which is to be restrained by the low conductance of the reverse-biased diode will flow through the capacitor (Fig. 3-13*b*).

3-8 CHARGE-CONTROL DESCRIPTION OF A DIODE

If the bias is in the forward direction, the potential barrier at the junction is lowered and holes from the *p* side enter the *n* side. Similarly, electrons from the *n* side move into the *p* side. This process of *minority-carrier injection* is discussed in Sec. 2-11. The excess hole density falls off exponentially with distance, as indicated in Fig. 3-14*a*. The shaded area under this curve is proportional to the injected charge.

For simplicity of discussion we assume that one side of the diode, say, the *p* material, is so heavily doped in comparison with the *n* side that the current I is carried across the junction entirely by holes moving from the *p* to the *n* side, or $I = I_{pn}(0)$. The excess minority charge Q will then exist only on the *n* side, and is given by the shaded area in the *n* region of Fig. 3-14*a* multiplied by the diode cross section A and the electronic charge q. Hence, from Eq. (2-47),

$$Q = \int_0^\infty Aqp'(0)\epsilon^{-x/L_p}\,dx = AqL_pp'(0) \tag{3-24}$$

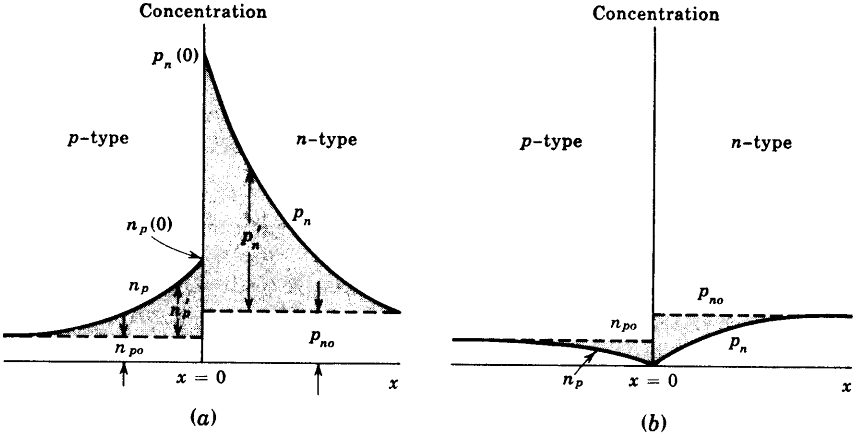


Fig. 3-14 Minority-carrier density distribution as a function of the distance x from a junction. (a) A forward-biased junction; (b) a reverse-biased junction. The excess hole (electron) density $p_n' = p_n - p_{no}$ ($n_p' = n_p - n_{po}$) is positive in (a) and negative in (b). (The transition region is assumed to be so small relative to the diffusion length that it is not indicated in this figure.)

The hole current I is given by $I_p(x)$ in Eq. (2-48), with $x = 0$, or

$$I = \frac{AqD_p p'(0)}{L_p} \tag{3-25}$$

Eliminating $p'(0)$ from Eqs. (3-24) and (3-25) yields

$$I = \frac{Q}{\tau} \tag{3-26}$$

where $\tau \equiv L_p^2/D_p \equiv \tau_p$ = mean lifetime for holes [Eq. (2-44)].

Equation (3-26) is an important relationship, referred to as the *charge-control description of a diode.* It states that the diode current (which consists of holes crossing the junction from the p to the n side) is proportional to the stored charge Q of excess minority carriers. The factor of proportionality is the reciprocal of the decay time constant (the mean lifetime τ) of the minority carriers. Thus, in the steady state, *the current I supplies minority carriers at the rate at which these carriers are disappearing because of the process of recombination.*

Charge Storage under Reverse Bias When an external voltage reverse-biases the junction, the steady-state density of minority carriers is as shown in Fig. 3-14*b*. Far from the junction the minority carriers are equal to their thermal-equilibrium values p_{no} and n_{po}, as is also the situation in Fig. 3-14*a*. As the minority carriers approach the junction they are rapidly swept across, and the density of minority carriers diminishes to zero at this junction. This result follows from the law of the junction, Eq. (3-4), since the concentration $p_n(0)$ reduces to zero for a negative junction potential V.

The injected charge under reverse bias is given by the shaded area in Fig. 3-14*b*. This charge is negative since it represents less charge than is available under conditions of thermal equilibrium with no applied voltage. From Eq. (3-26) with Q negative, the diode current I is negative and, of course, equals the reverse saturation current I_o in magnitude.

The charge-control characterization of a diode describes the device in terms of the current I and the stored charge Q, whereas the equivalent-circuit characterization uses the current I and the junction voltage V. One immediately apparent advantage of this charge-control description is that the exponential relationship between I and V is replaced by the linear dependence of I on Q. The charge Q also makes a simple parameter, the sign of which determines whether the diode is forward- or reverse-biased. The diode is forward-biased if Q is positive and reverse-biased if Q is negative.

3-9 DIFFUSION CAPACITANCE

For a forward bias a capacitance which is much larger than the transition capacitance C_T considered in Sec. 3-8 comes into play. The origin of this larger capacitance lies in the injected charge stored near the junction outside the transition region (Fig. 3-14*a*). It is convenient to introduce an incremental capacitance, defined as the rate of change of injected charge with voltage, called the *diffusion*, or *storage, capacitance* C_D.

Static Derivation of C_D We now make a quantitative study of C_D. From Eqs. (3-26) and (3-13)

$$C_D \equiv \frac{dQ}{dV} = \tau \frac{dI}{dV} = \tau g = \frac{\tau}{r} \tag{3-27}$$

where the diode incremental conductance $g \equiv dI/dV$. Substituting the expression for the diode incremental resistance $r = 1/g$ given in Eq. (3-14) into Eq. (3-27) yields

$$C_D = \frac{\tau I}{\eta V_T} \tag{3-28}$$

We see that *the diffusion capacitance is proportional to the current I*. In the derivation above we have assumed that the diode current I is due to holes only. If this assumption is not satisfied, Eq. (3-27) gives the diffusion capacitance C_{D_p} due to holes only, and a similar expression can be obtained for the diffusion capacitance C_{D_n} due to electrons. The total diffusion capacitance can then be obtained as the sum of C_{D_p} and C_{D_n} (Prob. 3-24).

For a reverse bias, g is very small and C_D may be neglected compared with C_T. For a forward current, on the other hand, C_D is usually much larger than C_T. For example, for germanium ($\eta = 1$) at $I = 26$ mA, $g = 1$ ℧, and $C_D = \tau$. If, say, $\tau = 20\ \mu$s, then $C_D = 20\ \mu$F, a value which is about a million times larger than the transition capacitance.

Despite the large value of C_D, the time constant rC_D (which is of importance in circuit applications) may not be excessive because the dynamic forward resistance $r = 1/g$ is small. From Eq. (3-27),

$$rC_D = \tau \tag{3-29}$$

Hence the diode time constant equals the mean lifetime of minority carriers, which lies in range of nanoseconds to hundreds of microseconds. The importance of τ in circuit applications is considered in the following section.

Diffusion Capacitance for an Arbitrary Input Consider the buildup of injected charge across a junction as a function of time when the potential is changed. In Fig. 3-15, the curve marked (1) is the steady-state value of p_n' for an applied voltage V. If the voltage at time t is increased by dV in the interval dt, then p_n' changes to that indicated by the curve marked (2) at time $t + dt$. The increase in charge dQ' in the time dt is proportional to the heavily shaded area in Fig. 3-15. Note that the concentration near the junction has increased markedly, whereas p_n far away from the junction has changed very little, because it takes time for the holes to diffuse into the n region. If the applied voltage is maintained constant at $V + dV$, the stored charge will continue to increase. Finally, at $t = \infty$, p_n follows the curve marked (3) [which is given by Eq. (3-4), with V replaced by $V + dV$]. The steady-state injected charge dQ, due to the increase in voltage by dV, is proportional to the total shaded area in Fig. 3-15. Clearly, $dQ - dQ'$ is represented by the lightly shaded area, and hence $dQ > dQ'$.

Since dQ' is the charge injected across the junction in time dt, the current is given by

$$i = \frac{dQ'}{dt} = C_D' \frac{dV}{dt} \tag{3-30}$$

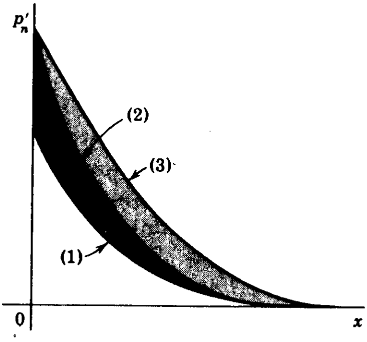


Fig. 3-15 The transient buildup of stored excess charge. The curve marked (1) gives the steady-state value of p_n' at a time t when the voltage is V. If the voltage increases by dV and held at $V + dV$, then p_n' is given by (2) at time $t + dt$ and by (3) at $t = \infty$.

where we define the *small-signal diffusion capacitance* C'_D by $C'_D \equiv dQ'/dV$. Note that the diode current is *not* given by the steady-state charge Q or static capacitance C_D.

$$i \neq \frac{dQ}{dt} \qquad \text{or} \qquad i \neq C_D \frac{dV}{dt} \tag{3-31}$$

Since $dQ' < dQ$, then $C'_D < C_D$.

From the above argument we conclude that the dynamic diffusion capacitance C'_D depends upon how the input voltage varies with time. To find C'_D, the equation of continuity must be solved for the given voltage waveform. This equation controls how p_n varies both as a function of x and t, and from $p_n(x, t)$ we can obtain the current. If the input varies with time in an arbitrary way, it may not be possible to define the diffusion capacitance in a unique manner.

Diffusion Capacitance for a Sinusoidal Input For the special case where the excitation varies sinusoidally with time, C'_D may be obtained from a solution of the equation of continuity. This analysis is carried out in Sec. 19-12, and C'_D is found to be a function of frequency. At low frequencies

$$C'_D = \tfrac{1}{2}\tau g \qquad \text{if } \omega\tau \ll 1 \tag{3-32}$$

which is half the value found in Eq. (3-27), based upon static considerations. For high frequencies, C'_D decreases with increasing frequency and is given by

$$C'_D = \left(\frac{\tau}{2\omega}\right)^{\frac{1}{2}} g \qquad \text{if } \omega\tau \gg 1 \tag{3-33}$$

3-10 JUNCTION–DIODE SWITCHING TIMES

When a diode is driven from the reversed condition to the forward state or in the opposite direction, the diode response is accompanied by a transient, and an interval of time elapses before the diode recovers to its steady state. The forward recovery time t_{fr} is the time difference between the 10 percent point of the diode voltage and the time when this voltage reaches and remains within 10 percent of its final value. It turns out that t_{fr} does not usually constitute a serious practical problem, and hence we here consider only the more important situation of reverse recovery.

Diode Reverse Recovery Time When an external voltage forward-biases a *p-n* junction, the steady-state density of minority carriers is as shown in Fig. 3-14*a*. The number of minority carriers is very large. These minority carriers have, in each case, been supplied from the other side of the junction, where, being majority carriers, they are in plentiful supply.

If the external voltage is suddenly reversed in a diode circuit which has

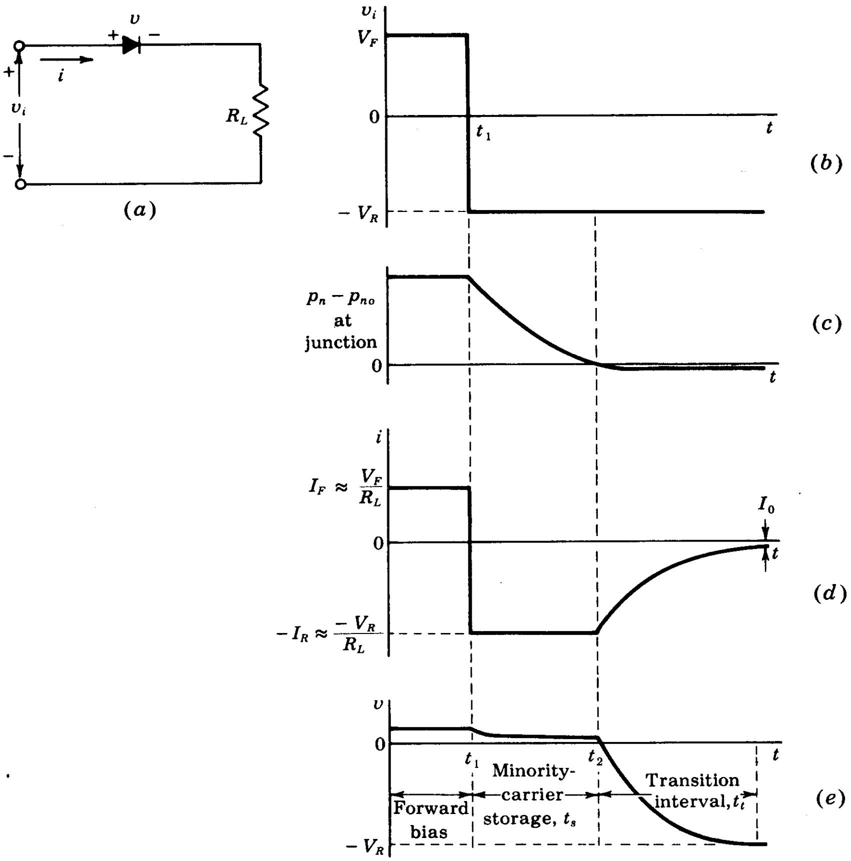


Fig. 3-16 The waveform in (b) is applied to the diode circuit in (a); (c) the excess carrier density at the junction; (d) the diode current; (e) the diode voltage.

been carrying current in the forward direction, the diode current will not immediately fall to its steady-state reverse-voltage value. For the current cannot attain its steady-state value until the minority-carrier distribution, which at the moment of voltage reversal had the form in Fig. 3-14*a*, reduces to the distribution in Fig. 3-14*b*. Until such time as the *injected*, or *excess*, *minority-carrier density* $p_n - p_{no}$ (or $n_p - n_{po}$) has dropped nominally to zero, the diode will continue to conduct easily, and the current will be determined by the external resistance in the diode circuit.

Storage and Transition Times The sequence of events which accompanies the reverse biasing of a conducting diode is indicated in Fig. 3-16. We consider that the voltage in Fig. 3-16*b* is applied to the diode-resistor circuit in Fig. 3-16*a*. For a long time, and up to the time t_1, the voltage $v_i = V_F$ has been in the direction to forward-bias the diode. The resistance R_L is assumed large enough so that the drop across R_L is large in comparison with the drop across the diode. Then the current is $i \approx V_F/R_L \equiv I_F$. At the time $t = t_1$ the input voltage reverses abruptly to the value $v = -V_R$. For the reasons described above, the current does not drop to zero, but instead reverses and remains at the value $i \approx -V_R/R_L \equiv -I_R$ until the time $t = t_2$. At $t = t_2$, as is seen in Fig. 3-16*c*, the minority-carrier density p_n at $x = 0$ has reached its equilibrium state p_{no}. If the diode ohmic resistance is R_d, then at the time t_1 the diode voltage falls slightly [by $(I_F + I_R)R_d$] but does not reverse. At $t = t_2$, when the excess minority carriers in the immediate neighborhood of the junction have been swept back across the junction, the diode voltage begins to reverse and the magnitude of the diode current begins to decrease. The interval t_1 to t_2, for the stored-minority charge to become zero, is called the *storage time* t_s.

The time which elapses between t_2 and the time when the diode has nominally recovered is called the *transition time* t_t. This recovery interval will be completed when the minority carriers which are at some distance from the junction have diffused to the junction and crossed it and when, in addition, the junction transition capacitance across the reverse-biased junction has charged through R_L to the voltage $-V_R$.

Manufacturers normally specify the reverse recovery time of a diode t_{rr} in a typical operating condition in terms of the current waveform of Fig. 3-16*d*. The time t_{rr} is the interval from the current reversal at $t = t_1$ until the diode has recovered to a specified extent in terms either of the diode current or of the diode resistance. If the specified value of R_L is larger than several hundred ohms, ordinarily the manufacturers will specify the capacitance C_L shunting R_L in the measuring circuit which is used to determine t_{rr}. Thus we find, for the Fairchild 1N3071, that with $I_F = 30$ mA and $I_R = 30$ mA, the time required for the reverse current to fall to 1.0 mA is 50 nsec. Again we find, for the same diode, that with $I_F = 30$ mA, $-V_R = -35$ V, $R_L = 2$ K, and $C_L = 10$ pF ($-I_R = -35/2 = -17.5$ mA), the time required for the diode to recover to the extent that its resistance becomes 400 K is $t_{rr} = 400$ nsec. Commercial switching-type diodes are available with times t_{rr} in the range from less than a nanosecond up to as high as 1 μs in diodes intended for switching large currents.

3-11 BREAKDOWN DIODES[3]

The reverse-voltage characteristic of a semiconductor diode, including the breakdown region, is redrawn in Fig. 3-17*a*. Diodes which are designed with adequate power-dissipation capabilities to operate in the breakdown region may be employed as voltage-reference or constant-voltage devices. Such diodes are known as *avalanche*, *breakdown*, or *Zener diodes*. They are used

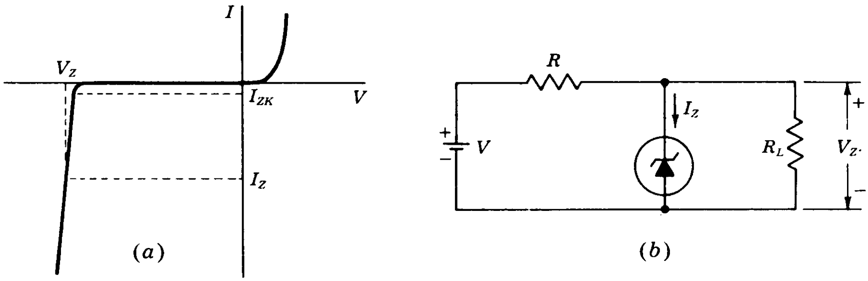


Fig. 3-17 (a) The volt-ampere characteristic of an avalanche, or Zener, diode. (b) A circuit in which such a diode is used to regulate the voltage across R_L against changes due to variations in load current and supply voltage.

characteristically in the manner indicated in Fig. 3-17*b*. The source V and resistor R are selected so that, initially, the diode is operating in the breakdown region. Here the diode voltage, which is also the voltage across the load R_L, is V_Z, as in Fig. 3-17*a*, and the diode current is I_Z. The diode will now regulate the load voltage against variations in load current and against variations in supply voltage V because, in the breakdown region, large changes in diode current produce only small changes in diode voltage. Moreover, as load current or supply voltage changes, the diode current will accommodate itself to these changes to maintain a nearly constant load voltage. The diode will continue to regulate until the circuit operation requires the diode current to fall to I_{ZK}, in the neighborhood of the knee of the diode volt-ampere curve. The upper limit on diode current is determined by the power-dissipation rating of the diode.

Avalanche Multiplication Two mechanisms of diode breakdown for increasing reverse voltage are recognized. Consider the following situation: A thermally generated carrier (part of the reverse saturation current) falls down the junction barrier and acquires energy from the applied potential. This carrier collides with a crystal ion and imparts sufficient energy to disrupt a covalent bond. In addition to the original carrier, a new electron-hole pair has now been generated. These carriers may also pick up sufficient energy from the applied field, collide with another crystal ion, and create still another electron-hole pair. Thus each new carrier may, in turn, produce additional carriers through collision and the action of disrupting bonds. This cumulative process is referred to as *avalanche multiplication*. It results in large reverse currents, and the diode is said to be in the region of *avalanche breakdown*.

Zener Breakdown Even if the initially available carriers do not acquire sufficient energy to disrupt bonds, it is possible to initiate breakdown through a direct rupture of the bonds. Because of the existence of the electric field at the junction, a sufficiently strong force may be exerted on a bound electron by the field to tear it out of its covalent bond. The new hole-electron pair which is created increases the reverse current. Note that this process, called *Zener breakdown,* does not involve collisions of carriers with the crystal ions (as does avalanche multiplication).

The field intensity $\mathcal{E}$ increases as the impurity concentration increases, for a fixed applied voltage (Prob. 3-25). It is found that Zener breakdown occurs at a field of approximately 2×10^7 V/m. This value is reached at voltages below about 6 V for heavily doped diodes. For lightly doped diodes the breakdown voltage is higher, and avalanche multiplication is the predominant effect. Nevertheless, the term *Zener* is commonly used for the *avalanche,* or *breakdown, diode* even at higher voltages. Silicon diodes operated in avalanche breakdown are available with maintaining voltages from several volts to several hundred volts and with power ratings up to 50 W.

Temperature Characteristics A matter of interest in connection with Zener diodes, as with semiconductor devices generally, is their temperature sensitivity. The temperature coefficient is given as the percentage change in reference voltage per centigrade degree change in diode temperature. These data are supplied by the manufacturer. The coefficient may be either positive or negative and will normally be in the range ±0.1 percent/°C. If the reference voltage is above 6 V, where the physical mechanism involved is avalanche multiplication, the temperature coefficient is positive. However, below 6 V, where true Zener breakdown is involved, the temperature coefficient is negative.

A qualitative explanation of the sign (positive or negative) of the temperature coefficient of V_Z is now given. A junction having a narrow depletion-layer width, and hence high field intensity, will break down by the Zener mechanism. An increase in temperature increases the energies of the valence electrons, and hence makes it easier for these electrons to escape from the covalent bonds. Less applied voltage is therefore required to pull these electrons from their positions in the crystal lattice and convert them into conduction electrons. Thus the Zener breakdown voltage decreases with temperature.

A junction with a broad depletion layer, and therefore a low field intensity, will break down by the avalanche mechanism. In this case we rely on intrinsic carriers to collide with valence electrons and create avalanche multiplication. As the temperature increases, the vibrational displacement of atoms in the crystal grows. This vibration increases the probability of collisions with the lattice atoms of the intrinsic particles as they cross the depletion width. The intrinsic holes and electrons thus have less of an opportunity to gain sufficient energy between collisions to start the avalanche process. Therefore the value of the avalanche voltage must increase with increased temperature.

Dynamic Resistance and Capacitance A matter of importance in connection with Zener diodes is the slope of the diode volt-ampere curve in the operating range. If the reciprocal slope $\Delta V_Z/\Delta I_Z$, called the *dynamic resistance,* is r, then a change ΔI_Z in the operating current of the diode produces a change $\Delta V_Z = r\,\Delta I_Z$ in the operating voltage. Ideally, $r = 0$, corresponding to a volt-ampere curve which, in the breakdown region, is precisely vertical. The variation of r at various currents for a series of avalanche diodes of fixed power-dissipation rating and various voltages show a rather broad minimum in the range 6 to 10 V. This minimum value of r is of the order of magnitude of a few ohms. However, for values of V_Z below 6 V or above 10 V, and particularly for small currents ($\sim$1 mA), r may be of the order of hundreds of ohms.

Some manufacturers specify the minimum current I_{ZK} (Fig. 3-17*a*) below which the diode should not be used. Since this current is on the knee of the above curve, where the dynamic resistance is large, then for currents lower than I_{ZK} the regulation will be poor. Some diodes exhibit a very sharp knee even down into the microampere region.

The capacitance across a breakdown diode is the transition capacitance, and hence varies inversely as some power of the voltage. Since C_T is proportional to the cross-sectional area of the diode, high-power avalanche diodes have very large capacitances. Values of C_T from 10 to 10,000 pF are common.

Additional Reference Diodes Zener diodes are available with voltages as low as about 2 V. Below this voltage it is customary, for reference and regulating purposes, to use diodes in the *forward* direction. As appears in Fig. 3-7, the volt-ampere characteristic of a forward-biased diode (sometimes called a *stabistor*) is not unlike the reverse characteristic, with the exception that, in the forward direction, the knee of the characteristic occurs at lower voltage. A number of forward-biased diodes may be operated in series to reach higher voltages. Such series combinations, packaged as single units, are available with voltages up to about 5 V, and may be preferred to reverse-biased Zener diodes, which at low voltages have very large values of dynamic resistance.

When it is important that a Zener diode operate with a low temperature coefficient, it may be feasible to operate an appropriate diode at a current where the temperature coefficient is at or near zero. Quite frequently, such operation is not convenient, particularly at higher voltages and when the diode must operate over a range of currents. Under these circumstances temperature-compensated avalanche diodes find application. Such diodes consist of a reverse-biased Zener diode with a positive temperature coefficient, combined in a single package with a forward-biased diode whose temperature coefficient is negative. As an example, the Transitron SV3176 silicon 8-V reference diode has a temperature coefficient of ± 0.001 percent/°C at 10 mA over the range -55 to $+100$°C. The dynamic resistance is only 1.5 Ω. The temperature coefficient remains below 0.002 percent/°C for currents in the range 8 to 12 mA. The voltage stability with time of some of these reference diodes is comparable with that of conventional standard cells.

When a high-voltage reference is required, it is usually advantageous (except of course with respect to economy) to use two or more diodes in series rather than a single diode. This combination will allow higher voltage, higher dissipation, lower temperature coefficient, and lower dynamic resistance.

3-12 THE TUNNEL DIODE

A *p-n* junction diode of the type discussed in Sec. 3-1 has an impurity concentration of about 1 part in 10^8. With this amount of doping, the width of the depletion layer, which constitutes a potential barrier at the junction, is of the order of a micron. This potential barrier restrains the flow of carriers from the side of the junction where they constitute majority carriers to the side where they constitute minority carriers. If the concentration of impurity atoms is greatly increased, say, to 1 part in 10^3 (corresponding to a density in excess of 10^{19} cm^{-3}), the device characteristics are completely changed. This new diode was announced in 1958 by Esaki,[4] who also gave the correct theoretical explanation for its volt-ampere characteristic.

The Tunneling Phenomenon The width of the junction barrier varies inversely as the square root of impurity concentration [Eq. (3-21)] and therefore is reduced to less than 100 Å (10^{-6} cm). This thickness is only about one-fiftieth the wavelength of visible light. Classically, a particle must have an energy at least equal to the height of a potential-energy barrier if it is to move from one side of the barrier to the other. However, for barriers as thin as those estimated above in the Esaki diode, the Schrödinger equation indicates that there is a large probability that an electron will penetrate *through* the barrier. This quantum-mechanical behavior is referred to as *tunneling*, and hence these high-impurity-density *p-n* junction devices are called *tunnel diodes*. The volt-ampere relationship is explained in Sec. 19-8 and is depicted in Fig. 3-18.

Characteristics of a Tunnel Diode[5] From Fig. 3-18 we see that the tunnel diode is an excellent conductor in the reverse direction (the *p* side of the junction negative with respect to the *n* side). Also, for small forward voltages (up to 50 mV for Ge), the resistance remains small (of the order of 5 Ω). At the *peak current* I_P corresponding to the voltage V_P, the slope dI/dV of the characteristic is zero. If V is increased beyond V_P, the current decreases. As a consequence, the dynamic conductance $g = dI/dV$ is negative. The tunnel diode exhibits a *negative-resistance characteristic* between the peak current I_P and the minimum value I_V, called the *valley current*. At the *valley voltage* V_V at which $I = I_V$, the conductance is again zero, and beyond this point the resistance becomes and remains positive. At the so-called *peak forward voltage* V_F the current again reaches the value I_P. For larger voltages the current increases beyond this value.

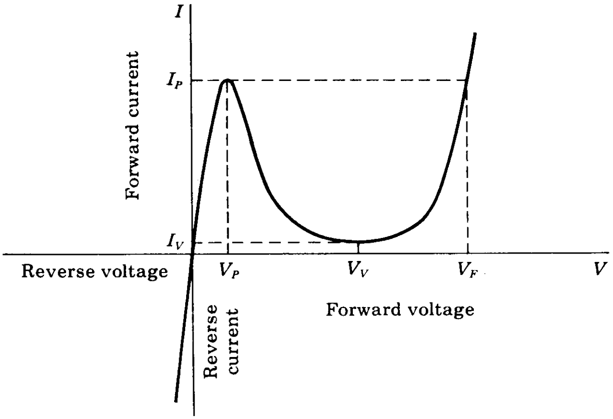


Fig. 3-18 Volt-ampere characteristic of a tunnel diode.

For currents whose values are between I_V and I_P, the curve is triple-valued, because each current can be obtained at three different applied voltages. It is this multivalued feature which makes the tunnel diode useful in pulse and digital circuitry.[6]

The standard circuit symbol for a tunnel diode is given in Fig. 3-19*a*. The small-signal model for operation in the negative-resistance region is indicated in Fig. 3-19*b*. The negative resistance $-R_n$ has a minimum at the point of inflection between I_P and I_V. The series resistance R_s is ohmic resistance. The series inductance L_s depends upon the lead length and the geometry of the dipole package. The junction capacitance C depends upon the bias and is usually measured at the valley point. Typical values for these parameters for a tunnel diode of peak current value $I_P = 10$ mA are $-R_n = -30\ \Omega$, $R_s = 1\ \Omega$, $L_s = 5$ nH, and $C = 20$ pF.

One interest in the tunnel diode is its application as a very high-speed switch. Since tunneling takes place at the speed of light, the transient response is limited only by total shunt capacitance (junction plus stray wiring capacitance) and peak driving current. Switching times of the order of a nanosecond are reasonable, and times as low as 50 ps have been obtained. A second application[5] of the tunnel diode is as a high-frequency (microwave) oscillator.

The most common commercially available tunnel diodes are made from germanium or gallium arsenide. It is difficult to manufacture a silicon tunnel

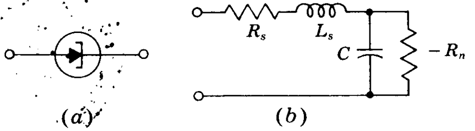


Fig. 3-19 (*a*) Symbol for a tunnel diode; (*b*) small-signal model in the negative-resistance region.

TABLE 3-1 Typical tunnel-diode parameters

	Ge	GaAs	Si
I_P/I_V	8	15	3.5
V_P, V	0.055	0.15	0.065
V_V, V	0.35	0.50	0.42
V_F, V	0.50	1.10	0.70

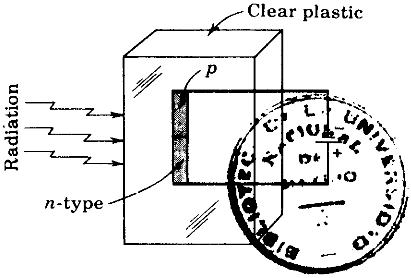
diode with a high ratio of peak-to-valley current I_P/I_V. Table 3-1 summarizes the important static characteristics of these devices. The voltage values in this table are determined principally by the particular semiconductor used and are almost independent of the current rating. Note that gallium arsenide has the highest ratio I_P/I_V and the largest voltage swing $V_F - V_P \approx 1.0$ V, as against 0.45 V for germanium. The peak current I_P is determined by the impurity concentration (the resistivity) and the junction area. For computer applications, devices with I_P in the range of 1 to 100 mA are most common. The peak point (V_P, I_P), which is in the tunneling region, is not a very sensitive function of temperature. However, the valley point (V_V, I_V), which is affected by the injection current, is quite temperature-sensitive.[5]

The advantages of the tunnel diode are low cost, low noise, simplicity, high speed, environmental immunity, and low power. The disadvantages of the diode are its low output-voltage swing and the fact that it is a two-terminal device. Because of the latter feature, there is no isolation between input and output, and this leads to serious circuit-design difficulties.

3-13 THE SEMICONDUCTOR PHOTODIODE

If a reverse-biased *p-n* junction is illuminated, the current varies almost linearly with the light flux. This effect is exploited in the semiconductor *photodiode*. This device consists of a *p-n* junction embedded in a clear plastic, as indicated in Fig. 3-20. Radiation is allowed to fall upon one surface across the junction. The remaining sides of the plastic are either painted black or enclosed in a metallic case. The entire unit is extremely small and has dimensions of the order of tenths of an inch.

Fig. 3-20 The construction of a semiconductor photodiode.



CIENCIAS

Volt-Ampere Characteristics If reverse voltages in excess of a few tenths of a volt are applied, an almost constant current (independent of the magnitude of the reverse bias) is obtained. The dark current corresponds to the reverse saturation current due to the thermally generated minority carriers. As explained in Sec. 3-2, these minority carriers "fall down" the potential hill at the junction, whereas this barrier does not allow majority carriers to cross the junction. Now if light falls upon the surface, additional electron-hole pairs are formed. In Sec. 2-8 we note that it is justifiable to consider the radiation solely as a *minority-carrier injector*. These injected minority carriers (for example, electrons in the p side) diffuse to the junction, cross it, and contribute to the current.

The reverse saturation current I_o in a p-n diode is proportional to the concentrations p_{no} and n_{po} of minority carriers in the n and p region, respectively. If we illuminate a reverse-biased p-n junction, the number of new hole-electron pairs is proportional to the number of incident photons. Hence the current under large reverse bias is $I = I_o + I_s$, where I_s, the short-circuit current, is proportional to the light intensity. Hence the volt-ampere characteristic is given by

$$I = I_s + I_o(1 - \epsilon^{V/\eta V_T}) \tag{3-34}$$

where I, I_s, and I_o represent the *magnitude* of the reverse current, and V is positive for a forward voltage and negative for a reverse bias. The parameter η is unity for germanium and 2 for silicon, and V_T is the volt equivalent of temperature defined by Eq. (3-10).

Typical photodiode volt-ampere characteristics are indicated in Fig. 3-21. The curves (with the exception of the dark-current curve) do not pass through the origin. The characteristics in the millivolt range and for positive bias are discussed in the following section, where we find that the photodiode may be used under either short-circuit or open-circuit conditions. It should be

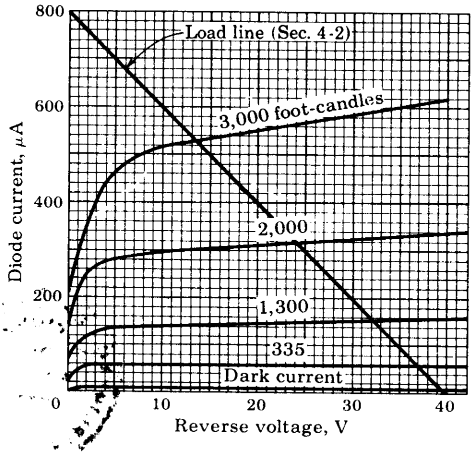


Fig. 3-21 Volt-ampere characteristics for the 1N77 germanium photodiode. (Courtesy of Sylvania Electric Products, Inc.)

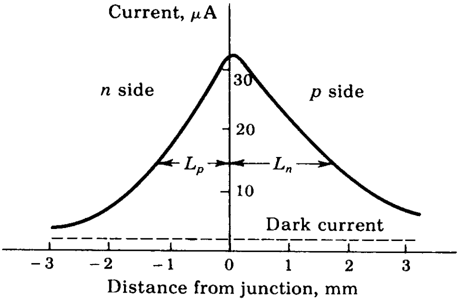


Fig. 3-22 Sensitivity of a semiconductor photodiode as a function of the distance of the light spot from the junction.

noted that the characteristics drift somewhat with age. The barrier capacitance $C_T \approx 10$ pF, the dynamic resistance $R \approx 50$ M, and the ohmic resistance $r \approx 100\ \Omega$.

Sensitivity with Position of Illumination The current in a reverse-biased semiconductor photodiode depends upon the diffusion of minority carriers to the junction. If the radiation is focused into a small spot far away from the junction, the injected minority carriers can recombine before diffusing to the junction. Hence a much smaller current will result than if the minority carriers were injected near the junction. The photocurrent as a function of the distance from the junction at which the light spot is focused is indicated in Fig. 3-22. The curve is somewhat asymmetrical because of the differences in the diffusion lengths of minority carriers in the *p* and *n* sides. Incidentally, the spectral response of the semiconductor photodiode is the same as that for a photoconductive cell, and is indicated in Fig. 2-12.

The *p-n* photodiode and, particularly, the improved *n-p-n* version described in Sec. 5-14 find extensive application in high-speed reading of computer punched cards and tapes, light-detection systems, reading of film sound track, light-operated switches, production-line counting of objects which interrupt a light beam, etc.

3-14 THE PHOTOVOLTAIC EFFECT[8]

In Fig. 3-21 we see that an almost constant reverse current due to injected minority carriers is collected in the *p-n* photodiode for large reverse voltages. If the applied voltage is reduced in magnitude, the barrier at the junction is reduced. This decrease in the potential hill does not affect the minority current (since these particles fall down the barrier), but when the hill is reduced sufficiently, some majority carriers can also cross the junction. These carriers correspond to a forward current, and hence such a flow will reduce the net (reverse) current. It is this increase in majority-carrier flow which accounts for the drop in the reverse current near the zero-voltage axis in Fig. 3-21.

An expanded view of the origin in this figure is indicated in Fig. 3-23. (Note that the first quadrant of Fig. 3-21 corresponds to the third quadrant of Fig. 3-23.)

The Photovoltaic Potential If a forward bias is applied, the potential barrier is lowered, and the majority current increases rapidly. When this majority current equals the minority current, the total current is reduced to zero. The voltage at which zero resultant current is obtained is called the *photovoltaic* potential. Since, certainly, no current flows under open-circuited conditions, the photovoltaic emf is obtained across the open terminals of a *p-n* junction.

An alternative (but of course equivalent) physical explanation of the photovoltaic effect is the following: In Sec. 3-1 we see that the height of the potential barrier at an open-circuited (nonilluminated) *p-n* junction adjusts itself so that the resultant current is zero, the electric field at the junction being in such a direction as to repel the majority carriers. If light falls on the surface, minority carriers are injected, and since these fall down the barrier, the minority current increases. Since under open-circuited conditions the total current must remain zero, the majority current (for example, the hole current in the *p* side) must increase the same amount as the minority current. This rise in majority current is possible only if the retarding field at the junction is reduced. Hence the barrier height is automatically lowered as a result of the radiation. Across the diode terminals there appears a voltage just equal to the amount by which the barrier potential is decreased. This potential is the photovoltaic emf and is of the order of magnitude of 0.5 V for a silicon and 0.1 V for a germanium cell.

The photovoltaic voltage $V_{\max}$ corresponds to an open-circuited diode. If $I = 0$ is substituted into Eq. (3-34), we obtain

$$V_{\max} = \eta V_T \ln \left(1 + \frac{I_s}{I_o}\right) \tag{3-35}$$

Since, except for very small light intensities, $I_s/I_o \gg 1$, then $V_{\max}$ increases logarithmically with I_s, and hence with illumination. Such a logarithmic relationship is obtained experimentally.

Maximum Output Power If a resistor R_L is placed directly across the diode terminals, the resulting current can be found at the intersection of the characteristic in Fig. 3-23 and the load line defined by $V = -IR_L$. If $R_L = 0$, then the output voltage V is zero, and for $R_L = \infty$, the output current I is zero. Hence, for these two extreme values of load, the output power is zero. If for each assumed value of R_L the values of V and I are read from Fig. 3-23 and $P = VI$ is plotted versus R_L, we can obtain the *optimum* load resistance to give maximum output power. For the types LS222 and LS223 photovoltaic light sensors, this optimum load is 3.4 K and $P_{\max} \approx 34\ \mu$W. When the *p-n*

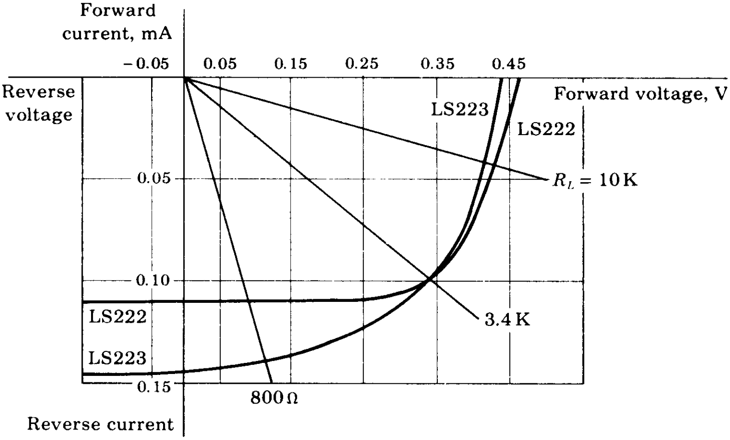


Fig. 3-23 Volt-ampere characteristics for the LS222 and LS223 *p-n* junction photodiodes at a light intensity of 500 fc. (Courtesy of Texas Instruments, Inc.)

photodiode is used as an energy converter (to transform radiant energy into electric energy), the optimum load resistance should be used.

The Short-circuit Current We see from Fig. 3-23 and Eq. (3-34) that a definite (nonzero) current is obtained for zero applied voltage. Hence a junction photocell can be used under short-circuit conditions. As already emphasized, this current I_s is proportional to the light intensity. Such a linear relationship is obtained experimentally.

Solar-energy Converters The current drain from a photovoltaic cell may be used to power electronic equipment or, more commonly, to charge auxiliary storage batteries. Such energy converters using sunlight as the primary energy are called *solar batteries* and are used in satellites like the Telstar. A silicon photovoltaic cell of excellent stability and high (~14 percent) conversion efficiency is made by diffusing a thin *n*-type impurity onto a *p*-type base. In direct noonday sunlight such a cell generates an open-circuit voltage of approximately 0.6 V.

3-15 LIGHT-EMITTING DIODES

Just as it takes energy to generate a hole-electron pair, so energy is released when an electron recombines with a hole. In silicon and germanium this recombination takes place through traps (Sec. 2-8) and the liberated energy goes into the crystal as heat. However, it is found that in other semiconductors, such as gallium arsenide, there is a considerable amount of direct recombination without the aid of traps. Under such circumstances the energy released when an electron falls from the conduction into the valence band appears in the form of radiation. Such a *p-n* diode is called a *light-emitting diode* (*LED*), although the radiation is principally in the infrared. The efficiency of the process of light generation increases with the injected current and with a decrease in temperature. The light is concentrated near the junction because most of the carriers are to be found within a diffusion length of the junction.

Under certain conditions, the emitted light is coherent (essentially monochromatic). Such a diode is called an *injection junction laser.*

REFERENCES

1. Gray, P. E., D. DeWitt, A. R. Boothroyd, and J. F. Gibbons: "Physical Electronics and Circuit Models of Transistors," vol. 2, Semiconductor Electronics Education Committee, John Wiley & Sons, Inc., New York, 1964.

   Shockley, W.: The Theory of *p-n* Junctions in Semiconductor and *p-n* Junction Transistors, *Bell System Tech. J.*, vol. 28, pp. 435–489, July, 1949.

   Middlebrook, R. D.: "An Introduction to Junction Transistor Theory," pp. 115–130, John Wiley & Sons, Inc., New York, 1957.

2. Phillips, A. B.: "Transistor Engineering," pp. 129–133, McGraw-Hill Book Company, New York, 1962.

   Sah, C. T.: Effect of Surface Recombination and Channel on P-N Junction and Transistor Characteristics, *IRE Trans. Electron. Devices*, vol. ED-9, no. 1, pp. 94–108, January, 1962.

3. Corning, J. J.: "Transistor Circuit Analysis and Design," pp. 40–42, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1965.

4. Esaki, L.: New Phenomenon in Narrow Ge *p-n* Junctions, *Phys. Rev.*, vol. 109, p. 603, 1958.

   Nanavati, R. P.: "Introduction to Semiconductor Electronics," chap. 12, McGraw-Hill Book Company, New York, 1963.

5. "Tunnel Diode Manual, TD-30," Radio Corporation of America, Semiconductor and Materials Division, Somerville, N.J., 1963.

   "Tunnel Diode Manual," General Electric Company, Semiconductor Products Dept., Liverpool, N.Y., 1961.

6. Millman, J., and H. Taub: "Pulse, Digital, and Switching Waveforms," chap. 13, McGraw-Hill Book Company, New York, 1965.

7. Shive, J. N.: "Semiconductor Devices," chaps. 8 and 9, D. Van Nostrand Company, Inc., Princeton, N.J., 1959.

8. Rappaport, R.: The Photovoltaic Effect and Its Utilization, *RCA Rev.*, vol. 21, no. 3, pp. 373–397, September, 1959.

Loferski, J. J.: Recent Research on Photovoltaic Solar Energy Converters, *Proc. IEEE*, vol. 51, no. 5, pp. 667–674, May, 1963.

Loferski, J. J., and J. J. Wysocki: Spectral Response of Photovoltaic Cells, *RCA Rev.*, vol. 22, no. 1, pp. 38–56, March, 1961.

REVIEW QUESTIONS

3-1 Consider an open-circuited *p-n* junction. Sketch curves as a function of distance across the junction of space charge, electric field, and potential.

3-2 (*a*) What is the order of magnitude of the space-charge width at a *p-n* junction? (*b*) What does this space charge consist of—electrons, holes, neutral donors, neutral acceptors, ionized donors, ionized acceptors, etc.?

3-3 (*a*) For a reverse-biased diode, does the transition region increase or decrease in width? (*b*) What happens to the junction potential?

3-4 Explain why the *p-n* junction contact potential *cannot* be measured by placing a voltmeter across the diode terminals.

3-5 Explain physically why a *p-n* diode acts as a rectifier.

3-6 (*a*) Write the *law of the junction.* (*b*) Define all terms in this equation. (*c*) What does this equation state for a large forward bias? (*d*) A large reverse bias?

3-7 Plot the minority-carrier current components and the total current in a *p-n* diode as a function of the distance from the junction.

3-8 Plot the hole current, the electron current, and the total current as a function of distance on both sides of a *p-n* junction. Indicate the transition region.

3-9 (*a*) Write the volt-ampere equation for a *p-n* diode. (*b*) Explain the meaning of each symbol.

3-10 Plot the volt-ampere curves for germanium and silicon to the same scale, showing the cutin value for each.

3-11 (*a*) How does the reverse saturation current of a *p-n* diode vary with temperature? (*b*) How does the diode voltage (at constant current) vary with temperature?

3-12 How does the dynamic resistance r of a diode vary with (*a*) current and (*b*) temperature? (*c*) What is the order of magnitude of r for silicon at room temperature and for a dc current of 1 mA?

3-13 (*a*) Sketch the piecewise linear characteristic of a diode. (*b*) What are the approximate cutin voltages for silicon and germanium?

3-14 Consider a step-graded *p-n* junction with equal doping on both sides of the junction ($N_A = N_D$). Sketch the charge density, field intensity, and potential as a function of distance from the junction for a reverse bias.

3-15 (*a*) How does the transition capacitance C_T vary with the depletion-layer width? (*b*) With the applied reverse voltage? (*c*) What is the order of magnitude of C_T?

3-16 What is a *varactor diode?*

3-17 Plot the minority-carrier concentration as a function of distance from a *p-n* junction in the *n* side only for (*a*) a forward-biased junction, (*b*) a negatively biased junction. Indicate the excess concentration and note where it is positive and where negative.

3-18 Under steady-state conditions the diode current is proportional to a charge Q. (*a*) What is the physical meaning of the factor of proportionality? (*b*) What charge does Q represent—transition layer charge, injected minority-carrier charge, majority-carrier charge, etc.?

3-19 (*a*) How does the diffusion capacitance C_D vary with dc diode current? (*b*) What does the product of C_D and the dynamic resistance of a diode equal?

3-20 What is meant by the *minority-carrier storage time* of a diode?

3-21 A diode in series with a resistor R_L is forward-biased by a voltage V_F. After a steady state is reached, the input changes to $-V_R$. Sketch the current as a function of time. Explain qualitatively the shape of this curve.

3-22 (*a*) Draw the volt-ampere characteristic of an avalanche diode. (*b*) What is meant by the *knee* of the curve? (*c*) By the dynamic resistance? (*d*) By the temperature coefficient?

3-23 Describe the physical mechanism for avalanche breakdown.

3-24 Describe the physical mechanism for Zener breakdown.

3-25 Draw a circuit which uses a breakdown diode to regulate the voltage across a load.

3-26 Sketch the volt-ampere characteristic of a tunnel diode. Indicate the negative-resistance portion.

3-27 Draw the small-signal model of the tunnel diode operating in the negative-resistance region. Define each circuit element.

3-28 (*a*) Draw the volt-ampere characteristics of a *p-n* photodiode. (*b*) Does the current correspond to a forward- or reverse-biased diode? (*c*) Each curve is drawn for a different value of what physical parameter?

3-29 (*a*) Write the equation for the volt-ampere characteristic of a photodiode. (*b*) Define each symbol in the equation.

3-30 (*a*) Sketch the curve of photodiode current as a function of the position of a narrow light source from the junction. (*b*) Explain the shape of the curve.

3-31 (*a*) Define *photovoltaic potential.* (*b*) What is its order of magnitude? (*c*) Does it correspond to a forward or a reverse voltage?

3-32 Explain how to obtain maximum power output from a photovoltaic cell.

3-33 What is a *light-emitting diode?*

4 / DIODE CIRCUITS

The *p-n* junction diode is considered as a circuit element. The concept of "load line" is introduced. The piecewise linear diode model is exploited in the following applications: clippers (single-ended and double-ended), comparators, diode gates, and rectifiers. Half-wave, full-wave, bridge, and voltage-doubling rectification are considered. Capacitor filters are discussed.

Throughout this chapter we shall assume that the input waveforms vary slowly enough so that the diode switching times (Sec. 3-10) may be neglected.

4-1 THE DIODE AS A CIRCUIT ELEMENT

The basic diode circuit, indicated in Fig. 4-1, consists of the device in series with a load resistance R_L and an input-signal source v_i. This circuit is now analyzed to find the instantaneous current i and the instantaneous diode voltage v, when the instantaneous input voltage is v_i.

The Load Line From Kirchhoff's voltage law (KVL),

$$v = v_i - iR_L \tag{4-1}$$

where R_L is the magnitude of the load resistance. This one equation is not sufficient to determine the two unknowns v and i in this expression. However, a second relation between these two variables is given by the static characteristic of the diode (Fig. 3-7). In Fig. 4-2*a* is indicated the simultaneous solution of Eq. (4-1) and the diode characteristic. The straight line, which is represented by Eq. (4-1), is called the *load line*. The load line passes through the points $i = 0$,

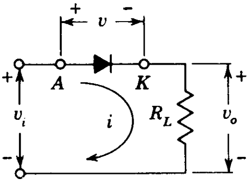


Fig. 4-1 The basic diode circuit. The anode (the p side) of the diode is marked A, and the cathode (the n side) is labeled K.

$v = v_i$, and $i = v_i/R_L$, $v = 0$. That is, the intercept with the voltage axis is v_i, and with the current axis is v_i/R_L. The slope of this line is determined, therefore, by R_L; the negative value of the slope is equal to $1/R_L$. The point of intersection A of the load line and the static curve gives the current i_A that will flow under these conditions. This construction determines the current in the circuit when the instantaneous input potential is v_i.

A slight complication may arise in drawing the load line because $i = v_i/R_L$ is too large to appear on the printed volt-ampere curve supplied by the manufacturer. Under such circumstance choose an arbitrary value of current I' which is on the vertical axis of the printed characteristic. Then the load line is drawn through the point P (Fig. 4-2a), where $i = I'$, $v = v_i - I'R_L$, and through a second point $i = 0$, $v = v_i$.

The Dynamic Characteristic Consider now that the input voltage is allowed to vary. Then the above procedure must be repeated for each voltage value. A plot of current vs. input voltage, called the *dynamic characteristic,* may be obtained as follows: The current i_A is plotted vertically above v_i at point B in Fig. 4-2b. As v_i changes, the slope of the load line does not

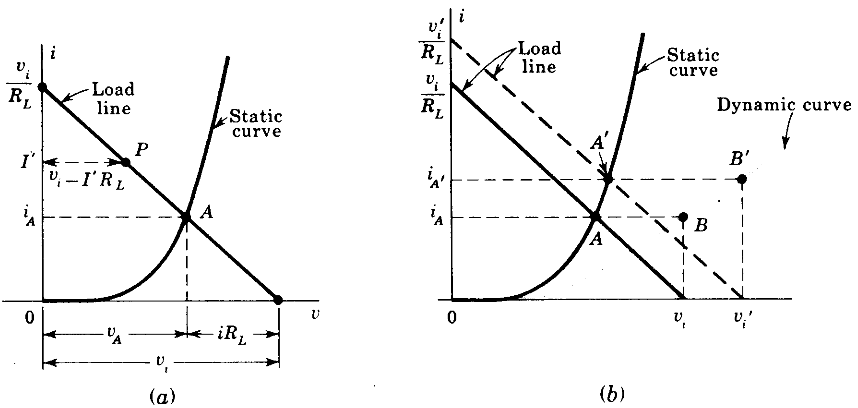


Fig. 4-2 (a) The intersection A of the load line with the diode static characteristic gives the current i_A corresponding to an instantaneous input voltage v_i. (b) The method of constructing the dynamic curve from the static curve and the load line.

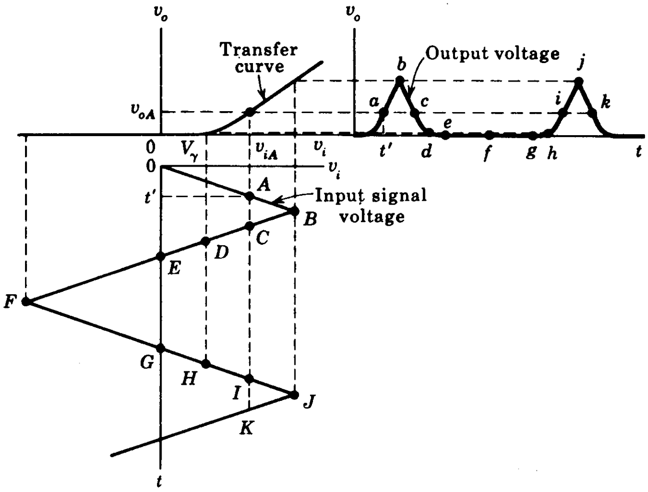


Fig. 4-3 The method of obtaining the output-voltage waveform from the transfer characteristic for a given input-signal-voltage waveform.

vary since R_L is fixed. Thus, when the applied potential has the value v_i', the corresponding current is $i_{A'}$. This current is plotted vertically above v_i' at B'. The resulting curve OBB' that is generated as v_i varies is the dynamic characteristic.

The Transfer Characteristic The curve which relates the output voltage v_o to the input v_i of any circuit is called the *transfer*, or *transmission, characteristic*. Since in Fig. 4-1 $v_o = iR_L$, then for this particular circuit the transfer curve has the same shape as the dynamic characteristic.

It must be emphasized that, regardless of the shape of the static volt-ampere characteristic or the waveform of the input signal, the resultant output waveshape can always be found graphically (at low frequencies) from the transfer curve. This construction is illustrated in Fig. 4-3. The input-signal waveform (not necessarily triangular) is drawn with its time axis vertically downward, so that the voltage axis is horizontal. Suppose that the input voltage has the value v_{iA} indicated by the point A at an instant t'. The corresponding output voltage is obtained by drawing a vertical line through A and noting the voltage v_{oA} where this line intersects the transfer curve. This value of v_o is then plotted (a) at an instant of time equal to t'. Similarly, points $b, c, d, \ldots$ of the output waveform correspond to points $B, C, D, \ldots$ of the input-voltage waveform. Note that $v_o = 0$ for $v_i < V_\gamma$, so that the diode acts as a *clipper* and a portion of the input signal does not appear at the output. Also note the distortion (the deviation from linearity) introduced into the

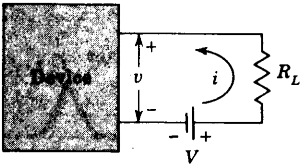


Fig. 4-4 **The output circuit of most devices consist of a supply voltage V in series with a load resistance R_L.**

output in the neighborhood of $v_i = V_\gamma$ because of the nonlinearity in the transfer curve in this region.

4-2 THE LOAD–LINE CONCEPT

We now show that the use of the load-line construction allows the graphical analysis of many circuits involving devices which are much more complicated than the *p-n* diode. The external circuit at the output of almost all devices consists of a dc (constant) supply voltage V in series with a load resistance R_L, as indicated in Fig. 4-4. Since KVL applied to this output circuit yields

$$v = V - iR_L \tag{4-2}$$

we once again have a straight-line relationship between output current i and output (device) voltage v. The load line passes through the point $i = 0$, $v = V$ and has a slope equal to $-1/R_L$ *independently of the device characteristics.* A *p-n* junction diode or an avalanche diode possesses a single volt-ampere characteristic at a given temperature. However, most other devices must be described by a family of curves. For example, refer to Fig. 3-21, which gives the volt-ampere characteristics of a photodiode, where a separate curve is drawn, for each fixed value of light intensity. The load line superimposed upon these characteristics corresponds to a 40-V supply and a load resistance of 40/800 M = 50 K. Note that, from the intersection of the load line with the curve for an intensity $L = 3{,}000$ fc, we obtain a photodiode current of 530 μA and a device voltage of 13.5 V. For $L = 2{,}000$ fc, $i = 320$ μA, and $v = 24.0$ V, etc.

The volt-ampere characteristics of a transistor (which is discussed in the following chapter) are similar to those in Fig. 3-21 for the photodiode. However, the independent parameter, which is held constant for each curve, is the input transistor current instead of light intensity. The output circuit is identical with that in Fig. 4-4, and the graphical analysis begins with the construction of the load line.

4-3 THE PIECEWISE LINEAR DIODE MODEL

If the reverse resistance R_r is included in the diode characteristic of Fig. 3-9, the piecewise linear and continuous volt-ampere characteristic of Fig. 4-5*a* is

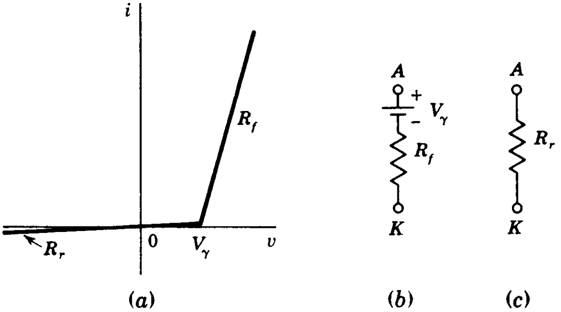


Fig. 4-5 (a) The piecewise linear volt-ampere characteristic of a *p-n* diode. (b) The large-signal model in the ON, or forward, direction (anode *A* more positive than V_γ with respect to the cathode). (c) The model in the OFF, or reverse, direction ($v < V_\gamma$).

obtained. The diode is a *binary* device, in the sense that it can exist in only one of two possible states; that is, the diode is either ON or OFF at a given time. If the voltage applied across the diode exceeds the cutin potential V_γ with the anode A (the p side) more positive than the cathode K (the n side), the diode is forward-biased and is said to be in the ON state. The large-signal model for the ON state is indicated in Fig. 4-5*b* as a battery V_γ in series with the low forward resistance R_f (of the order of a few tens of ohms or less). For a reverse bias ($v < V_\gamma$) the diode is said to be in its OFF state. The large-signal model for the OFF state is indicated in Fig. 4-5*c* as a large reverse resistance R_r (of the order of several hundred kilohms or more). Usually R_r is so much larger than any other resistance in the diode circuit that this reverse resistance may be considered to be infinite. We shall henceforth assume that $R_r = \infty$, unless otherwise stated.

A Simple Application Consider that in the basic diode circuit of Fig. 4-1 the input is sinusoidal, so that $v_i = V_m \sin \alpha$, where $\alpha = \omega t$, $\omega = 2\pi f$, and f is the frequency of the input excitation. Assume that the piecewise linear model of Fig. 4-5 (with $R_r = \infty$) is valid. The current in the forward direction ($v_i > V_\gamma$) may then be obtained from the equivalent circuit of Fig. 4-6*a*. We have

$$i = \frac{V_m \sin \alpha - V_\gamma}{R_L + R_f} \tag{4-3}$$

for $v_i = V_m \sin \alpha \geq V_\gamma$ and $i = 0$ for $v_i < V_\gamma$. This waveform is plotted in Fig. 4-6*b*, where the cutin angle ϕ is given by

$$\phi = \arcsin \frac{V_\gamma}{V_m} \tag{4-4}$$

If, for example, $V_m = 2V_\gamma$, then $\phi = 30°$. For silicon (germanium),

$$V_\gamma = 0.6 \text{ V } (0.2 \text{ V}),$$

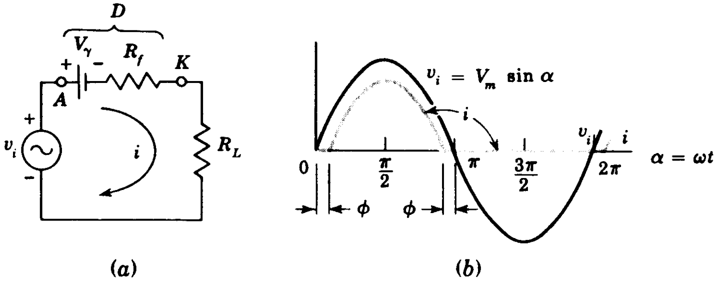


Fig. 4-6 (a) The equivalent circuit of a diode D (in the ON state) in series with a load resistance R_L and a sinusoidal voltage v_i. (b) The input waveform v_i and the rectified current i.

and hence a cutin angle of 30° is obtained for very small peak sinusoidal voltages; 1.2 V (0.4 V) for Si (Ge). On the other hand, if $V_m \geq 10$ V, then $\phi \leq 3.5°$ (1.2°) for Si (Ge) and the cutin angle may be neglected; the diode conducts essentially for a full half cycle. Such a rectifier is considered in more detail in Sec. 4-8.

Incidentally, the circuit of Fig. 4-6 may be used to charge a battery from an ac supply line. The battery V_B is placed in series with the diode D, and R_L is adjusted to supply the desired dc (average) charging current. The instantaneous current is given by Eq. (4-3), with V_B added to V_γ.

The Break Region The piecewise linear approximation given in Fig. 4-5*a* indicates an abrupt discontinuity in slope at V_γ. Actually, the transition of the diode from the OFF condition to the ON state is not abrupt. Therefore the waveform transmitted through a clipper or a rectifier will not show an abrupt change of attenuation at a break point, but instead there will exist a *break* region, that is, a region over which the slope of the diode characteristic changes gradually from a very small to a very large value. We shall now estimate the range of voltage of this break region.

The break point is defined at the voltage V_γ, where the diode resistance changes discontinuously from the very large value R_r to the very small value R_f. Hence, let us arbitrarily define the break region as the voltage change over which the diode resistance is multiplied by some large factor, say 100. The incremental resistance $r \equiv dV/dI = 1/g$ is, from Eq. (3-13),

$$r = \eta \frac{V_T}{I_o} \epsilon^{-V/\eta V_T} \qquad (4\text{-}5)$$

If $V_1(V_2)$ is the potential at which $r = r_1(r_2)$, then

$$\frac{r_1}{r_2} = \epsilon^{(V_2 - V_1)/\eta V_T} \qquad (4\text{-}6)$$

For $r_1/r_2 = 100$, $\Delta V \equiv V_2 - V_1 = \eta V_T \ln 100 = 0.12$ V for Ge ($\eta = 1$) and 0.24 V for Si ($\eta = 2$) at room temperature. Note that the break region ΔV is only one- or two-tenths of a volt. If the input signal is large compared with this small range, then the piecewise linear volt-ampere approximation and models of Fig. 4-5 are valid.

Analysis of Diode Circuits Using the Piecewise Linear Model Consider a circuit containing several diodes, resistors, supply voltages, and sources of excitation. A general method of analysis of such a circuit consists in assuming (guessing) the state of each diode. For the ON state, replace the diode by a battery V_γ in series with a forward resistance R_f, and for the OFF state replace the diode by the reverse resistance R_r (which can usually be taken as infinite), as indicated in Fig. 4-5*b* and *c*. After the diodes have been replaced by these piecewise linear models, the entire circuit is linear and the currents and voltages everywhere can be calculated using Kirchhoff's voltage and current laws. The assumption that a diode is ON can then be verified by observing the sign of the current through it. If the current is in the forward direction (from anode to cathode), the diode is indeed ON and the initial guess is justified. However, if the current is in the reverse direction (from cathode to anode), the assumption that the diode is ON has been proved incorrect. Under this circumstance the analysis must begin again with the diode assumed to be OFF.

Analogous to the above trial-and-error method, we test the assumption that a diode is OFF by finding the voltage across it. If this voltage is either in the reverse direction or in the forward direction but with a voltage less than V_γ, the diode is indeed OFF. However, if the diode voltage is in the forward direction and exceeds V_γ, the diode must be ON and the original assumption is incorrect. In this case the analysis must begin again by assuming the ON state for this diode.

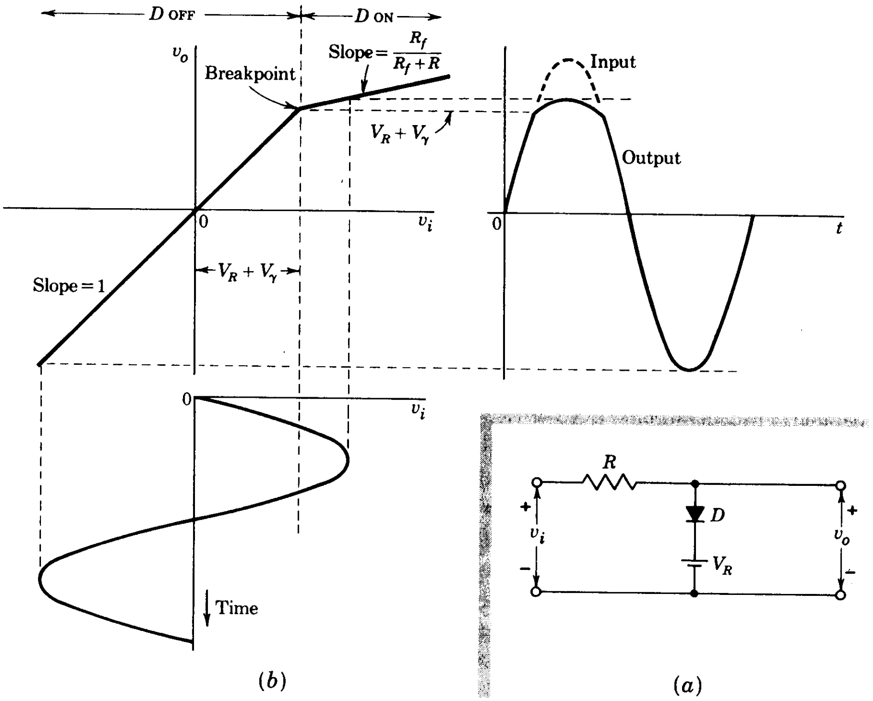
The above method of analysis will be employed in the study of the diode circuits which follows.

4-4 CLIPPING (LIMITING) CIRCUITS

Clipping circuits are used to select for transmission that part of an arbitrary waveform which lies above or below some reference level. Clipping circuits are also referred to as voltage (or current) *limiters*, *amplitude selectors*, or *slicers*.

In the above sense, Fig. 4-1 is a clipping circuit, and input voltages below V_γ are *not* transmitted to the output, as is evident from the waveforms of Figs. 4-3 and 4-6. Some of the more commonly employed clipping circuits are now to be described.

Consider the circuit of Fig. 4-7*a*. Using the piecewise linear model, the transfer characteristic of Fig. 4-7*b* is obtained, as may easily be verified. For example, if D is OFF, the diode voltage $v < V_\gamma$ and $v_i < V_\gamma + V_R$. How-



(b)

(a)

Fig. 4-7 (a) A diode clipping circuit which transmits that part of the waveform more negative than $V_R + V_\gamma$. (b) The piecewise linear transmission characteristic of the circuit. A sinusoidal input and the clipped output are shown.

ever, if D is OFF, there is no current in R and $v_o = v_i$. This argument justifies the linear portion (with slope unity) of the transmission characteristic extending from arbitrary negative values to $v_i = V_R + V_\gamma$. For v_i larger than $V_R + V_\gamma$, the diode conducts, and it behaves as a battery V_γ in series with a resistance R_f, so that increments Δv_i in the input are attenuated and appear at the output as increments $\Delta v_o = \Delta v_i R_f/(R_f + R)$. This verifies the linear portion of slope $R_f/(R_f + R)$ for $v_i > V_R + V_\gamma$ in the transfer curve. Note that the transmission characteristic is piecewise linear and continuous and has a break point at $V_R + V_\gamma$.

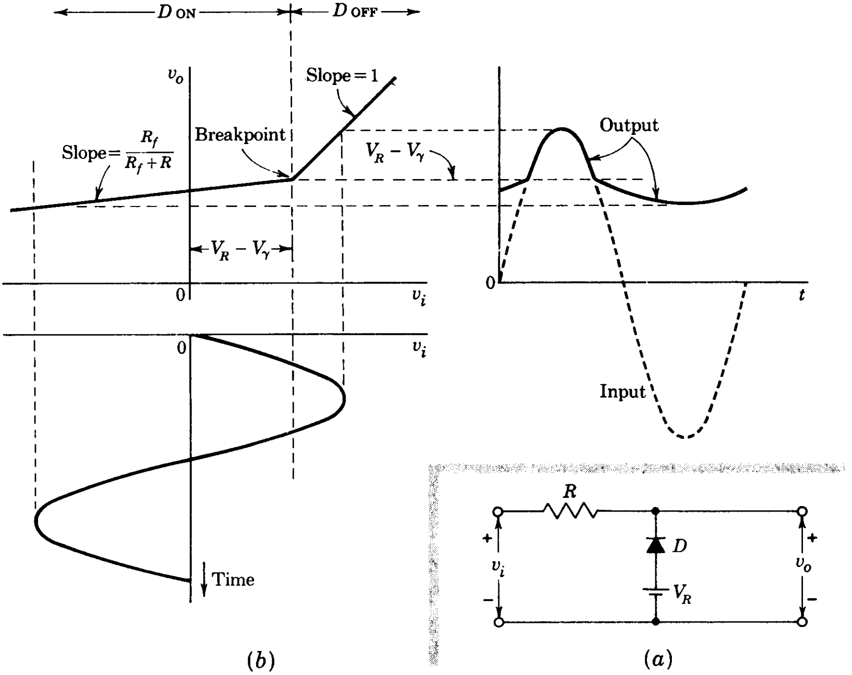
Figure 4-7*b* shows a sinusoidal input signal of amplitude large enough so that the signal makes excursions past the break point. The corresponding output exhibits a suppression of the positive peak of the signal. If $R_f \ll R$, this suppression will be very pronounced, and the positive excursion of the output will be sharply limited at the voltage $V_R + V_\gamma$. The output will appear as though the positive peak had been "clipped off" or "sliced off."

Often it turns out that $V_R \gg V_\gamma$, in which case one may consider that V_R itself is the limiting reference voltage.

In Fig. 4-8*a* the clipping circuit has been modified in that the diode in Fig. 4-7*a* has been reversed. The corresponding piecewise linear representation of the transfer characteristic is shown in Fig. 4-8*b*. In this circuit, the portion of the waveform more positive than $V_R - V_\gamma$ is transmitted without attenuation, but the less positive portion is greatly suppressed.

In Figs. 4-7*b* and 4-8*b* we have assumed R_r arbitrarily large in comparison with R. If this condition does not apply, the transmission characteristics must be modified. The portions of these curves which are indicated as having unity slope must instead be considered to have a slope $R_r/(R_r + R)$.

In a transmission region of a diode clipping circuit we require that $R_r \gg R$, for example, that $R_r = kR$, where k is a large number. In the attenuation region, we require that $R \gg R_f$, for example, that $R = kR_f$. From these two equations we deduce that $R = \sqrt{R_f R_r}$ and that $k = \sqrt{R_r/R_f}$. On this



(*b*) (*a*)

Fig. 4-8 (*a*) A diode clipping circuit which transmits that part of the waveform more positive than $V_R - V_\gamma$. (*b*) The piecewise linear transmission characteristic of the circuit. A sinusoidal input and the clipped output are shown.

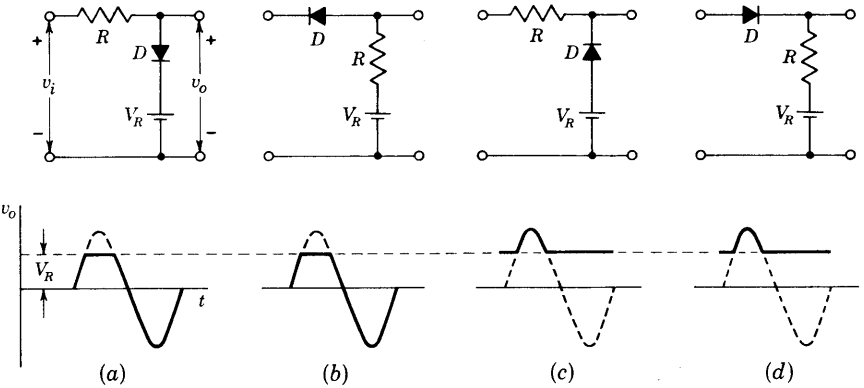


Fig. 4-9 Four diode clipping circuits. In (*a*) and (c) the diode appears as a shunt element. In (*b*) and (*d*) the diode appears as a series element. Under each circuit appears the output waveform (solid) for a sinusoidal input. The clipped portion of the input is shown dashed.

basis we conclude that it is reasonable to select R as the geometrical mean of R_r and R_f. And we note that the ratio R_r/R_f may well serve as a figure of merit for diodes used in the present application.

Additional Clipping Circuits Figures 4-7 and 4-8 appear again in Fig. 4-9, together with variations in which the diodes appear as series elements. If in each case a sinusoid is applied at the input, the waveforms at the output will appear as shown by the heavy lines. In these output waveforms we have neglected V_γ in comparison with V_R and we have assumed that the break region is negligible in comparison with the amplitude of the waveforms. We have also assumed that $R_r \gg R \gg R_f$. In two of these circuits the portion of the waveform transmitted is that part which lies below V_R; in the other two the portion above V_R is transmitted. In two the diode appears as an element in series with the signal lead; in two it appears as a shunt element. The use of the diode as a series element has the disadvantage that when the diode is OFF and it is intended that there be no transmission, fast signals or high-frequency waveforms may be transmitted to the output through the diode capacitance. The use of the diode as a shunt element has the disadvantage that when the diode is open (back-biased) and it is intended that there be transmission, the diode capacitance, together with all other capacitance in shunt with the output terminals, will round sharp edges of input waveforms and attenuate high-frequency signals. A second disadvantage of the use of the diode as a shunt element is that in such circuits the impedance R_s of the source which supplies V_R must be kept low. This requirement does not arise in circuits where V_R is in series with R, which is normally large compared with R_f.

4-5 CLIPPING AT TWO INDEPENDENT LEVELS

Diode clippers may be used in pairs to perform double-ended limiting at independent levels. A parallel, a series, or a series-parallel arrangement may be used. A parallel arrangement is shown in Fig. 4-10*a*. Figure 4-10*b* shows the piecewise linear and continuous input-output voltage curve for the circuit in Fig. 4-10*a*. The transfer curve has two break points, one at $v_o = v_i = V_{R1}$ and a second at $v_o = v_i = V_{R2}$, and has the following characteristics (assuming $V_{R2} > V_R \gg V_\gamma$ and $R_f \ll R$):

Input v_i	*Output* v_o	*Diode states*
$v_i \leq V_{R1}$	$v_o = V_{R1}$	$D1$ ON, $D2$ OFF
$V_{R1} < v_i < V_{R2}$	$v_o = v_i$	$D1$ OFF, $D2$ OFF
$v_i \geq V_{R2}$	$v_o = V_{R2}$	$D1$ OFF, $D2$ ON

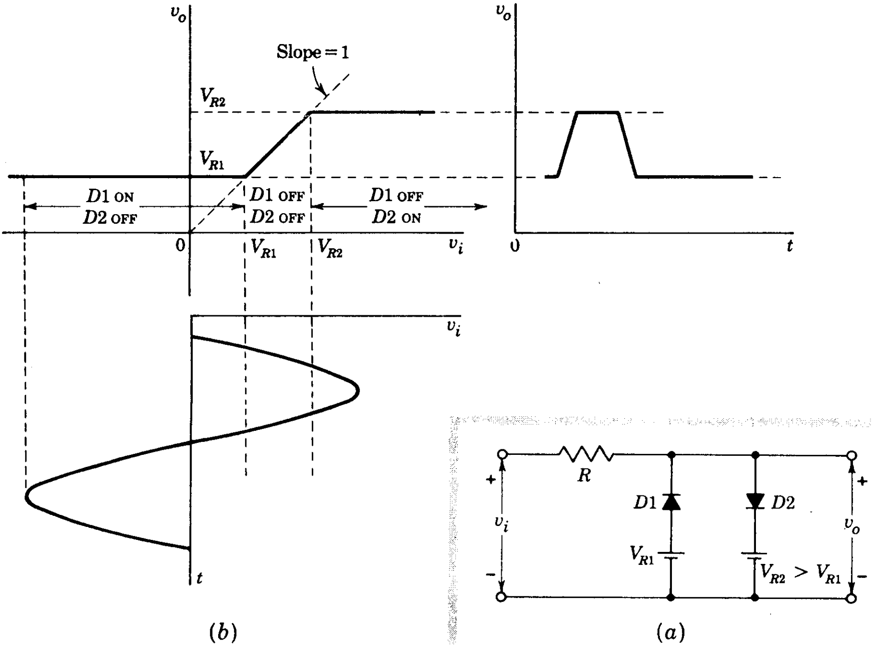


Fig. 4-10 (*a*) A double-diode clipper which limits at two independent levels. (*b*) The piecewise linear transfer curve for the circuit in (*a*). The doubly clipped output for a sinusoidal input is shown.

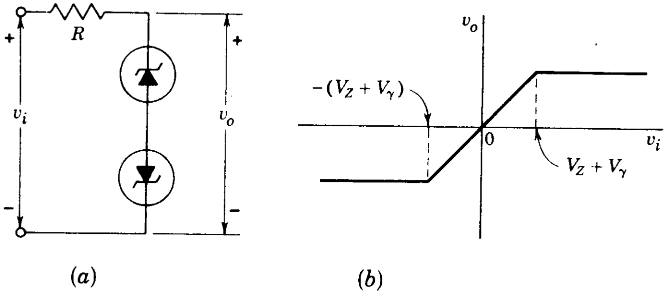


Fig. 4-11 (*a*) A double-ended clipper using avalanche diodes; (*b*) the transfer characteristic.

The circuit of Fig. 4-10*a* is referred to as a *slicer* because the output contains a slice of the input between the two reference levels V_{R1} and V_{R2}.

The circuit is used as a means of converting a sinusoidal waveform into a square wave. In this application, to generate a symmetrical square wave, V_{R1} and V_{R2} are adjusted to be numerically equal but of opposite sign. The transfer characteristic passes through the origin under these conditions, and the waveform is clipped symmetrically top and bottom. If the amplitude of the sinusoidal waveform is very large in comparison with the difference in the reference levels, the output waveform will have been *squared*.

Two avalanche diodes in series opposing, as indicated in Fig. 4-11*a*, constitute another form of double-ended clipper. If the diodes have identical characteristics, a symmetrical limiter is obtained. If the breakdown (Zener) voltage is V_Z and if the diode cutin voltage is V_γ, then the transfer characteristic of Fig. 4-11*b* is obtained.

Catching or Clamping Diodes Consider that v_i and R in Fig. 4-10*a* represent Thévenin's equivalent circuit at the output of a device, such as an amplifier. In other words, R is the output resistance and v_i is the open-circuit output signal. In such a situation $D1$ and $D2$ are called *catching diodes.* The reason for this terminology should be clear from Fig. 4-12, where we see

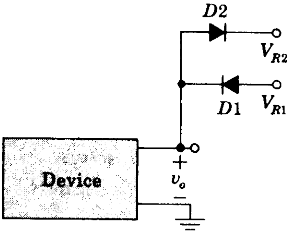


Fig. 4-12 Catching diodes $D1$ and $D2$ limit the output excursion of the device between V_{R1} and V_{R2}.

that $D1$ "catches" the output v_o and does not allow it to fall below V_{R1}, whereas $D2$ "catches" v_o and does not permit it to rise above V_{R2}.

Generally, whenever a node becomes connected through a low resistance (as through a conducting diode) to some reference voltage V_R, we say that the node has been clamped to V_R, since the voltage at that point in the circuit is unable to depart appreciably from V_R. In this sense the diodes in Fig. 4-12 are called *clamping diodes.*

A circuit for clamping the extremity of a periodic waveform to a reference voltage is considered in Sec. 4-11.

4-6 COMPARATORS

The nonlinear circuits which we have used to perform the operation of clipping may also be used to perform the operation of *comparison.* In this case the circuits become elements of a *comparator system* and are usually referred to simply as *comparators.* A comparator circuit is one which may be used to mark the instant when an arbitrary waveform attains some reference level. The distinction between comparator circuits and the clipping circuits considered earlier is that in a comparator there is no interest in reproducing any part of the signal waveform. For example, the comparator output may consist of an abrupt departure from some quiescent level which occurs at the time the signal attains the reference level but is otherwise independent of the signal. Or the comparator output may be a sharp pulse which occurs when signal and reference are equal.

The diode circuit of Fig. 4-13*a* which we encountered earlier as a clipping circuit is used here in a comparator operation. For the sake of illustration the input signal is taken as a ramp. This input crosses the voltage level $v_i = V_R + V_\gamma$ at time $t = t_1$. The output remains quiescent at $v_o \doteq V_R$ until $t = t_1$, after which it rises with the input signal.

The device to which the comparator output is applied will respond when the comparator voltage has risen to some level V_o above V_R. However, the precise voltage at which this device responds is subject to some variability Δv_o because of gradual changes which result from aging of components, temperature changes, etc. As a consequence (as shown in Fig. 4-13*b*) there will be a variability Δt in the precise moment at which this device responds and an uncertainty Δv_i in the input voltage corresponding to Δt. Furthermore, if the device responds in the range Δv_o, the device will respond not at $t = t_1$, but at some later time t_2. The situation may be improved by increasing the slope of the rising portion of the output waveform v_o. If the diode were indeed ideal, it would be advantageous to follow the comparator of Fig. 4-13*a* by an amplifier. However, because of the exponential characteristic of a physical diode, such an anticipated advantage is not realized.[1]

Although an amplifier which follows a diode-resistor comparator does not improve the sharpness of the comparator break, an amplifier *preceding* the

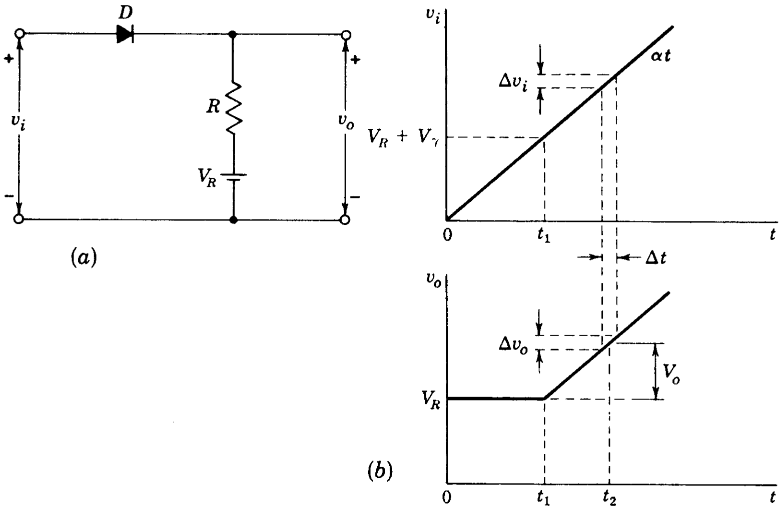


Fig. 4-13 (a) A diode comparator; (b) the comparison operation is illustrated with a ramp input signal v_i, and the corresponding output waveform is indicated.

comparator will do so. Thus, suppose that the input signal to a diode comparator must go through a range Δv_i to carry the comparator through its uncertainty region. Then, if the amplifier has a gain A, the input signal need only go through the range $\Delta v_i/A$ to carry the comparator output through the same voltage range. The amplifier must be direct-coupled and must be extremely stable against drift due to aging of components, temperature change, etc. Such an amplifier is the difference or operational amplifier discussed in Sec. 15-2. Comparators are treated in detail in Sec. 16-11.

4-7 SAMPLING GATE

An ideal sampling gate is a transmission circuit in which the output is an exact reproduction of an input waveform during a selected time interval and is zero otherwise. The time interval for transmission is selected by an externally impressed signal, called the *control*, or *gating, signal*, and is usually rectangular in shape. These sampling gates are also referred to as *transmission gates*, or *time-selection circuits*.

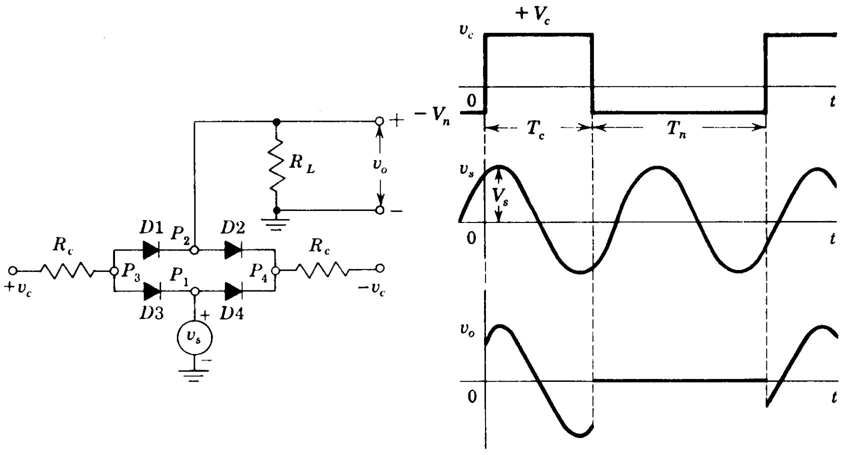
A four-diode sampling gate is indicated in Fig. 4-14a. This circuit has the topology of a bridge with the external signal v_s applied at node P_1, the output v_o taken across the load R_L at node P_2, and symmetrical control voltages $+v_c$ and $-v_c$ applied to nodes P_3 and P_4 through the control resistors R_c. The rectangularly shaped v_c, the sinusoidal v_s (it could be of arbitrary waveshape), and the sampled output v_o are drawn in Fig. 4-14*b*. Note that the period of v_c need not be the same as that of v_s, although in most practical systems the period of v_c would equal or be an integral multiple of that of v_s.

If we assume ideal diodes with $V_\gamma = 0$, $R_f = 0$, $R_r = \infty$, the operation of the circuit is easily understood. During the time interval T_c, when $v_c = V_c$, all four diodes conduct and the voltage across each is zero. Hence nodes P_1 and P_2 are at the same potential and $v_o = v_s$. The output is therefore an exact replica of the input during the selection time T_c. During the time T_n, when $v_c = -V_n$, all four diodes are nonconducting and the current in R_L is zero, so that $v_o = 0$.

We must now justify the statements made in the preceding paragraph that during T_c all diodes conduct and during T_n all diodes are nonconducting. Consider the situation when $v_c = -V_n$, $-v_c = +V_n$, and $v_s = V_s =$ the (positive) peak signal voltage. Let us assume that all diodes are reverse-biased, so that $v_o = 0$, and then justify this assumption (Sec. 4-3). From Fig. 4-14*a* we see that $D1$ and $D2$ are each reverse-biased by V_n, that $D3$ is reverse-biased by $V_n + V_s$, and that the voltage across $D4$ (in the forward direction) is $V_s - V_n$. Hence diodes $D1$, $D2$, and $D3$ are OFF for any value of V_n or V_s and $D4$ is OFF provided that $V_n \geq V_s$, or the minimum value of V_n is given by

$$(V_n)_{\min} = V_s \tag{4-7}$$

In other words, there is a restriction on the control-gate amplitude during the nonconducting interval T_n; the minimum value of V_n just equals the peak



(*a*) (*b*)

Fig. 4-14 (a) A four-diode-bridge sampling gate. (b) The control v_c, the signal v_s, and the output v_o waveforms.

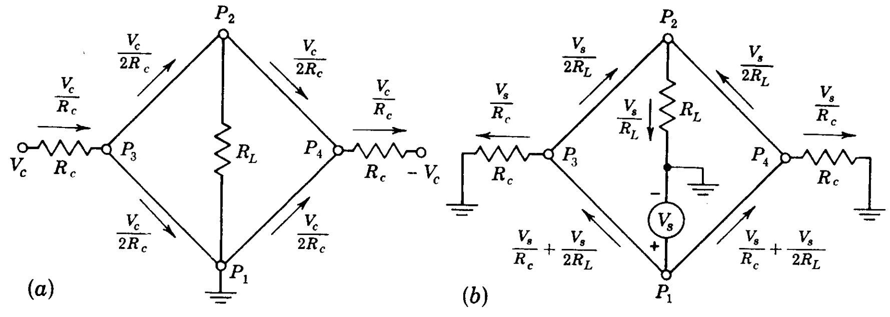


Fig. 4-15 The diodes in Fig. 4-14 are replaced by short circuits. (a) The currents due to V_c; (b) the currents due to V_s.

signal voltage V_s if we require that all four diodes be nonconducting during this interval.

Consider now the situation during T_c, when $v_c = +V_c$, $-v_c = -V_c$, and $v_s = V_s$. We now assume that all four diodes are ON and then determine the restriction required upon V_c so that each diode current is indeed in the forward direction. The current in each diode consists of two components, one due to V_c (as indicated in Fig. 14-15a) and the other due to V_s (as indicated in Fig. 14-15b). The current due to V_c is $V_c/2R_c$ and is in the forward direction in each diode, but the current due to V_s is in the reverse direction in $D3$ (between P_3 and P_1) and in $D2$ (between P_2 and P_4). The larger reverse current is in $D3$ and equals $V_s/R_c + V_s/2R_L$, and hence this quantity must be less than $V_c/2R_c$. The minimum value of V_c is therefore given by

$$(V_c)_{\text{min}} = V_s\left(2 + \frac{R_c}{R_L}\right) \tag{4-8}$$

Balance Conditions Assume that $v_s = 0$ but that the four diodes are not identical in the parameters V_γ and R_f (which are now no longer taken to be zero). Then the bridge will not be balanced and node P_2 is not at the same potential (ground) as P_1. Under these circumstances a portion of the rectangular control waveform appears at the output. In other words, during T_c the output is $v_o = V_c'$, instead of zero. If now the restriction $v_s = 0$ is removed, the sampled portion of the output v_o in Fig. 4-14b will be raised with respect to ground by the voltage V_c', and v_o is said to be "sitting upon a pedestal." Fortunately, all four diodes can be fabricated simultaneously on a tiny chip of silicon by integrated-circuit techniques (Chap. 7), and this ensures matched diodes, so that the pedestal is minimized. It must be emphasized that even with identical diodes a pedestal will exist in the output if the two control waveforms are not balanced (one control signal must be the negative of the other). Other sampling circuits which minimize control-signal imbalance are possible.[2]

4-8 RECTIFIERS

Almost all electronic circuits require a dc source of power. For portable low-power systems batteries may be used. More frequently, however, electronic equipment is energized by a *power supply*, a piece of equipment which converts the alternating waveform from the power lines into an essentially direct voltage. The study of ac-to-dc conversion is initiated in this section.

A Half-wave Rectifier A device, such as the semiconductor diode, which is capable of converting a sinusoidal input waveform (whose average value is zero) into a unidirectional (though not constant) waveform, with a nonzero average component, is called a *rectifier*. The basic circuit for half-wave rectification is shown in Fig. 4-16. Since in a rectifier circuit the input $v_i = V_m \sin \omega t$ has a peak value V_m which is very large compared with the cutin voltage V_γ of the diode, we assume in the following discussion that $V_\gamma = 0$. (The condition $V_\gamma \neq 0$ is treated in Sec. 4-3, and the current waveform is shown in Fig. 4-6*b*.) With the diode idealized to be a resistance R_f in the ON state and an open circuit in the OFF state, the current i in the diode or load R_L is given by

$$\begin{aligned} i &= I_m \sin \alpha && \text{if } 0 \leq \alpha \leq \pi \\ i &= 0 && \text{if } \pi \leq \alpha \leq 2\pi \end{aligned} \tag{4-9}$$

where $\alpha \equiv \omega t$ and

$$I_m \equiv \frac{V_m}{R_f + R_L} \tag{4-10}$$

The transformer secondary voltage v_i is shown in Fig. 4-16*b*, and the rectified

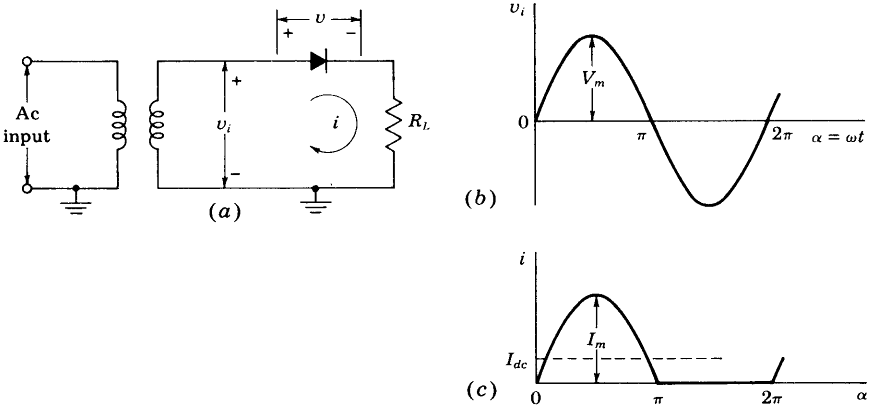


Fig. 4-16 (a) Basic circuit of half-wave rectifier. (b) Transformer sinusoidal secondary voltage v_i. (c) Diode and load current i.

current in Fig. 4-16*c*. Note that the output current is unidirectional. We now calculate this nonzero value of the average current.

A dc ammeter is constructed so that the needle deflection indicates the average value of the current passing through it. By definition, the average value of a periodic function is given by the area of one cycle of the curve divided by the base. Expressed mathematically,

$$I_{\text{dc}} = \frac{1}{2\pi}\int_0^{2\pi} i\,d\alpha \tag{4-11}$$

For the half-wave circuit under consideration, it follows from Eqs. (4-9) that

$$I_{\text{dc}} = \frac{1}{2\pi}\int_0^{\pi} I_m \sin\alpha\,d\alpha = \frac{I_m}{\pi} \tag{4-12}$$

Note that the upper limit of the integral has been changed from 2π to π since the instantaneous current in the interval from π to 2π is zero and so contributes nothing to the integral.

The Diode Voltage The dc output voltage is clearly given as

$$V_{\text{dc}} = I_{\text{dc}}R_L = \frac{I_m R_L}{\pi} \tag{4-13}$$

However, the reading of a dc voltmeter placed across the diode is *not* given by $I_{\text{dc}}R_f$ because the diode cannot be modeled as a constant resistance, but rather it has two values: R_f in the ON state and ∞ in the OFF state.

A dc voltmeter reads the average value of the voltage across its terminals. Hence, to obtain V'_{dc} across the diode, the instantaneous voltage must be plotted as in Fig. 4-17 and the average value obtained by integration. Thus

$$\begin{aligned} V'_{\text{dc}} &= \frac{1}{2\pi}\left(\int_0^{\pi} I_m R_f \sin\alpha\,d\alpha + \int_{\pi}^{2\pi} V_m \sin\alpha\,d\alpha\right) \\ &= \frac{1}{\pi}(I_m R_f - V_m) = \frac{1}{\pi}[I_m R_f - I_m(R_f + R_L)] \end{aligned}$$

where use has been made of Eq. (4-10). Hence

$$V'_{\text{dc}} = -\frac{I_m R_L}{\pi} \tag{4-14}$$

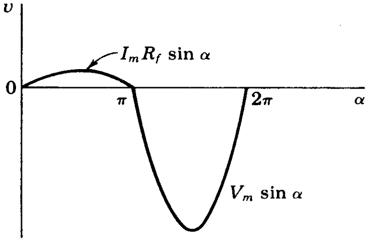


Fig. 4-17 The voltage across the diode in Fig. 4-16.

This result is negative, which means that if the voltmeter is to read upscale, its positive terminal must be connected to the cathode of the diode. From Eq. (4-13) the dc diode voltage is seen to be equal to the negative of the dc voltage across the load resistor. This result is evidently correct because the sum of the dc voltages around the complete circuit must add up to zero.

The AC Current (Voltage) *A root-mean-square ammeter (voltmeter) is constructed so that the needle deflection indicates the effective, or rms, current (voltage).* Such a "square-law" instrument may be of the thermocouple type. By definition, the effective or rms value squared of a periodic function of time is given by the area of one cycle of the curve, which represents the square of the function, divided by the base. Expressed mathematically,

$$I_{\mathrm{rms}} = \left(\frac{1}{2\pi}\int_0^{2\pi} i^2\,d\alpha\right)^{\frac{1}{2}} \tag{4-15}$$

By use of Eqs. (4-9), it follows that

$$I_{\mathrm{rms}} = \left(\frac{1}{2\pi}\int_0^{\pi} I_m{}^2\sin^2\alpha\,d\alpha\right)^{\frac{1}{2}} = \frac{I_m}{2} \tag{4-16}$$

Applying Eq. (4-15) to the *sinusoidal input voltage,* we obtain

$$V_{\mathrm{rms}} = \frac{V_m}{\sqrt{2}} \tag{4-17}$$

Most ac meters are of the rectifier type discussed in Sec. 4-9, instead of being true rms reading instruments.

Regulation The variation of dc output voltage as a function of dc load current is called *regulation.* The percentage regulation is defined as

$$\%\ \text{regulation} \equiv \frac{V_{\text{no load}} - V_{\text{load}}}{V_{\text{load}}} \times 100\% \tag{4-18}$$

where *no load* refers to zero current and *load* indicates the normal load current. For an ideal power supply the output voltage is independent of the load (the output current) and the percentage regulation is zero.

The variation of V_{dc} with I_{dc} for the half-wave rectifier is obtained as follows: From Eqs. (4-12) and (4-10),

$$I_{\mathrm{dc}} = \frac{I_m}{\pi} = \frac{V_m/\pi}{R_f + R_L} \tag{4-19}$$

Solving Eq. (4-19) for $V_{\mathrm{dc}} = I_{\mathrm{dc}}R_L$, we obtain

$$V_{\mathrm{dc}} = \frac{V_m}{\pi} - I_{\mathrm{dc}}R_f \tag{4-20}$$

This result is consistent with the circuit model given in Fig. 4-18 for the dc voltage and current. Note that the rectifier circuit functions as if it were

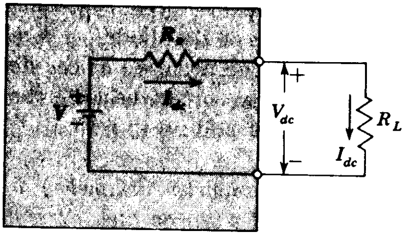


Fig. 4-18 The Thévenin's model which gives the dc voltage and current for a power supply. For the half-wave circuit of Fig. 4-16, $V = V_m/\pi$ and $R_o = R_f$. For the full-wave circuit of Fig. 4-19, $V = 2V_m/\pi$ and $R_o = R_f$. For the full-wave rectifier with a capacitor filter (Sec. 4-10), $V_o = V_m$ and $R_o = 1/4fC$.

a constant (open-circuit) voltage source $V = V_m/\pi$ in series with an effective internal resistance (the *output resistance*) $R_o = R_f$. This model shows that V_{dc} equals V_m/π at no load and that the dc voltage decreases linearly with an increase in dc output current. In practice, the resistance R_s of the transformer secondary is in series with the diode, and in Eq. (4-20) R_s should be added to R_f. The best method of estimating the diode resistance is to obtain a regulation plot of V_{dc} versus I_{dc} in the laboratory. The negative slope of the resulting straight line gives $R_f + R_s$.

Thévenin's Theorem This theorem states that *any two-terminal linear network may be replaced by a generator equal to the open-circuit voltage between the terminals in series with the output impedance seen at this port.* Clearly, Fig. 4-18 represents a Thévenin's model, and hence a rectifier behaves as a linear circuit with respect to average current and voltage.

A Full-wave Rectifier The circuit of a full-wave rectifier is shown in Fig. 4-19*a*. This circuit is seen to comprise two half-wave circuits so connected that conduction takes place through one diode during one half of the power cycle and through the other diode during the second half of the cycle.

The current to the load, which is the sum of these two currents, has the form shown in Fig. 4-19*b*. The dc and rms values of the load current and voltage in such a system are readily found to be

$$I_{\text{dc}} = \frac{2I_m}{\pi} \qquad I_{\text{rms}} = \frac{I_m}{\sqrt{2}} \qquad V_{\text{dc}} = \frac{2I_mR_L}{\pi} \tag{4-21}$$

where I_m is given by Eq. (4-10) and V_m is the peak transformer secondary voltage from one end to the center tap. Note by comparing Eq. (4-21) with Eq. (4-13) that the dc output voltage for the full-wave connection is twice that for the half-wave circuit.

From Eqs. (4-10) and (4-21) we find that the dc output voltage varies with current in the following manner:

$$V_{\text{dc}} = \frac{2V_m}{\pi} - I_{\text{dc}}R_f \tag{4-22}$$

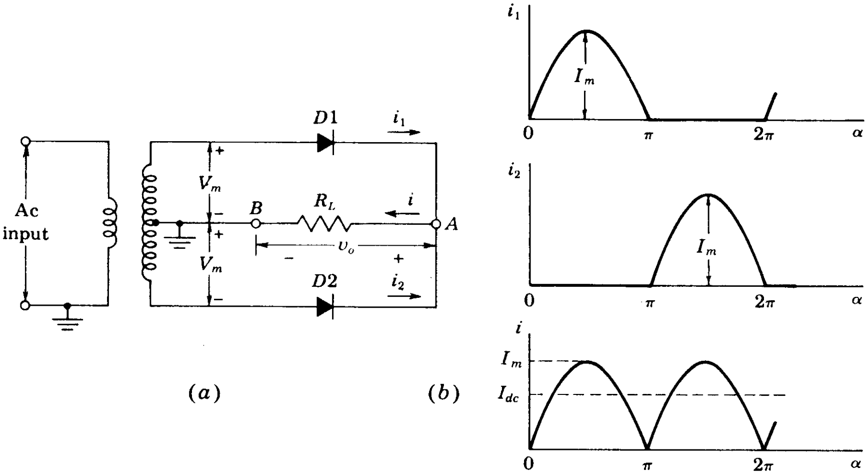


Fig. 4-19 (a) A full-wave rectifier circuit. (b) The individual diode currents and the load current i. The output voltage is $v_o = iR_L$.

This expression leads to Thévenin's dc model of Fig. 4-18, except that the internal (open-circuit) supply is $V = 2V_m/\pi$ instead of V_m/π.

Peak Inverse Voltage For each rectifier circuit there is a maximum voltage to which the diode can be subjected. This potential is called the *peak inverse voltage* because it occurs during that part of the cycle when the diode is nonconducting. From Fig. 4-16 it is clear that, for the half-wave rectifier, the peak inverse voltage is V_m. We now show that, for a full-wave circuit, twice this value is obtained. At the instant of time when the transformer secondary voltage to midpoint is at its peak value V_m, diode $D1$ is conducting and $D2$ is nonconducting. If we apply KVL around the outside loop and neglect the small voltage drop across $D1$, we obtain $2V_m$ for the peak inverse voltage across $D2$. Note that this result is obtained without reference to the nature of the load, which can be a pure resistance R_L or a combination of R_L and some reactive elements which may be introduced to "filter" the ripple. We conclude that, *in a full-wave circuit, independently of the filter used, the peak inverse voltage across each diode is twice the maximum transformer voltage measured from midpoint to either end.*

4-9 OTHER FULL-WAVE CIRCUITS

A variety of other rectifier circuits find extensive use. Among these are the bridge circuit, several voltage-doubling circuits, and a number of voltage-

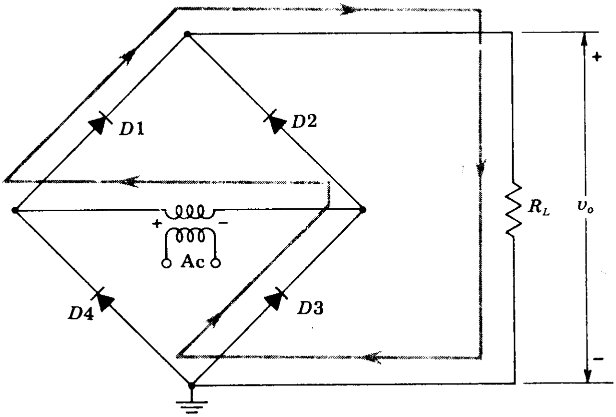


Fig. 4-20 Full-wave bridge circuit.

multiplying circuits. The bridge circuit finds application not only for power circuits, but also as a rectifying system in rectifier ac meters for use over a fairly wide range of frequencies.

The Bridge Rectifier The essentials of the bridge circuit are shown in Fig. 4-20. To understand the action of this circuit, it is necessary only to note that two diodes conduct simultaneously. For example, during the portion of the cycle when the transformer polarity is that indicated in Fig. 4-20, diodes 1 and 3 are conducting, and current passes from the positive to the negative end of the load. The conduction path is shown in the figure. During the next half cycle, the transformer voltage reverses its polarity, and diodes 2 and 4 send current through the load in the same direction as during the previous half cycle.

The principal features of the bridge circuit are the following: The currents drawn in both the primary and the secondary of the supply transformer are sinusoidal, and therefore a smaller transformer may be used than for the full-wave circuit of the same output; a transformer without a center tap is used; and each diode has only transformer voltage across it on the inverse cycle. The bridge circuit is thus suitable for high-voltage applications.

The Rectifier Meter This instrument, illustrated in Fig. 4-21, is essentially a bridge-rectifier system, except that no transformer is required. Instead, the voltage to be measured is applied through a multiplier resistor R to two corners of the bridge, a dc milliammeter being used as an indicating instrument across the other two corners. Since the dc milliammeter reads average values of current, the meter scale is calibrated to give rms values when a sinusoidal

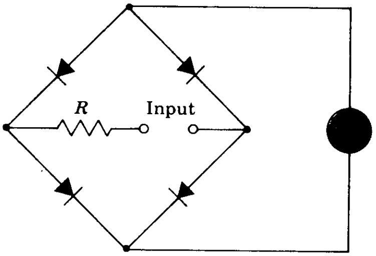


Fig. 4-21 The rectifier voltmeter.

voltage is applied to the input terminals. As a result, this instrument will not read correctly when used with waveforms which contain appreciable harmonics.

Voltage Multipliers A common voltage-doubling circuit which delivers a dc voltage approximately equal to twice the transformer maximum voltage at no load is shown in Fig. 4-22. This circuit is operated by alternately charging each of the two capacitors to the transformer peak voltage V_m, current being continually drained from the capacitors through the load. The capacitors also act to smooth out the ripple in the output.

This circuit is characterized by poor regulation unless very large capacitors are used. The inverse voltage across the diodes during the nonconducting cycle is twice the transformer peak voltage. The action of this circuit will be better understood after the capacitor filter is studied, in the following section.

4-10 CAPACITOR FILTERS

Filtering is frequently effected by shunting the load with a capacitor. The action of this system depends upon the fact that the capacitor stores energy during the conduction period and delivers this energy to the load during the inverse, or nonconducting, period. In this way, the time during which the current passes through the load is prolonged, and the ripple is considerably decreased. The ripple voltage is defined as the deviation of the load voltage from its average or dc value.

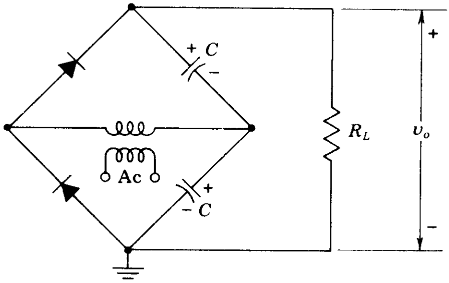


Fig. 4-22 The bridge voltage-doubling circuit. This is the single-phase full-wave bridge circuit of Fig. 4-19 with two capacitors replacing two diodes.

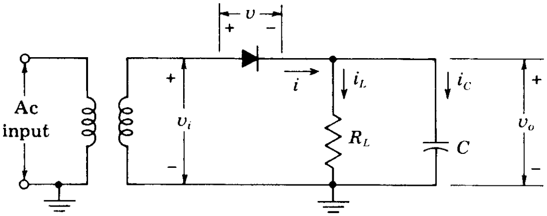


Fig. 4-23 A half-wave capacitor-filtered rectifier.

Consider the half-wave capacitive rectifier of Fig. 4-23. Suppose, first, that the load resistance $R_L = \infty$. The capacitor will charge to the potential V_m, the transformer maximum value. Further, the capacitor will maintain this potential, for no path exists by which this charge is permitted to leak off, since the diode will not pass a negative current. The diode resistance is infinite in the inverse direction, and no charge can flow during this portion of the cycle. Consequently, the filtering action is perfect, and the capacitor voltage v_o remains constant at its peak value, as is seen in Fig. 4-24.

The voltage v_o across the capacitor is, of course, the same as the voltage across the load resistor, since the two elements are in parallel. The diode voltage v is given by

$$v = v_i - v_o \tag{4-23}$$

We see from Fig. 4-24 that the diode voltage is always negative and that the peak inverse voltage is twice the transformer maximum. Hence the presence of the capacitor causes the peak inverse voltage to increase from a value equal to the transformer maximum when no capacitor filter is used to a value equal to twice the transformer maximum value when the filter is used.

Suppose, now, that the load resistor R_L is finite. Without the capacitor input filter, the load current and the load voltage during the conduction period will be sinusoidal functions of time. The inclusion of a capacitor in the circuit results in the capacitor charging in step with the applied voltage. Also, the capacitor must discharge through the load resistor, since the diode will prevent a current in the negative direction. Clearly, the diode acts as a switch which permits charge to flow into the capacitor when the transformer voltage exceeds the capacitor voltage, and then acts to disconnect the power source when the transformer voltage falls below that of the capacitor.

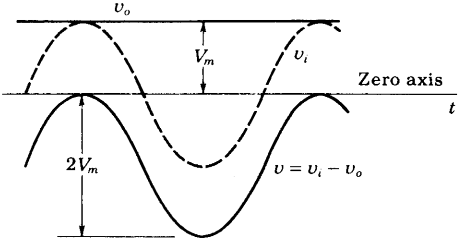


Fig. 4-24 Voltages in a half-wave capacitor-filtered rectifier at no load. The output voltage v_o is a constant, indicating perfect filtering. The diode voltage v is negative for all values of time, and the peak inverse voltage is $2V_m$.

The analysis now proceeds in two steps. First, the conditions during conduction are considered, and then the situation when the diode is nonconducting is investigated.

Diode Conducting If the diode drop is neglected, the transformer voltage is impressed directly across the load. Hence the output voltage is $v_o = V_m \sin \omega t$. The question immediately arises: Over what interval of time is this equation applicable? In other words, over what portion of each cycle does the diode remain conducting? The point at which the diode starts to conduct is called the *cutin point*, and that at which it stops conducting is called the *cutout point*. The cutout time t_1 and the cutin time t_2 are indicated in Fig. 4-25, where we observe that the output waveform consists of portions of sinusoids (when the diode is ON) joined to exponential segments (when the diode is OFF).

We now calculate the cutout angle by finding the expression for the diode current i and then noting when $i = 0$. When the diode conducts, $v_o = V_m \sin \omega t$ and i is the sum of the load resistor current i_L and the capacitor current i_C. Hence

$$i = \frac{v_o}{R_L} + C\frac{dv_o}{dt} = \frac{V_m}{R_L}\sin \omega t + \omega C V_m \cos \omega t \tag{4-24}$$

This current is of the form $i = I_m \sin(\omega t + \psi)$, where

$$I_m \equiv V_m \sqrt{\frac{1}{R_L{}^2} + \omega^2 C^2} \qquad \psi \equiv \arctan \omega C R_L \tag{4-25}$$

The cutout time t_1 is found by setting $i = 0$ at $t = t_1$, which leads to

$$\omega t_1 = \pi - \psi \tag{4-26}$$

for the cutout angle ωt_1 in the first cycle. The diode current is indicated in Fig. 4-25.

Equation (4-25) shows that the use of a large capacitance to improve the filtering at a given load R_L is accompanied by a high-peak diode current I_m. For a specified average load current, i becomes more peaked and the conduc-

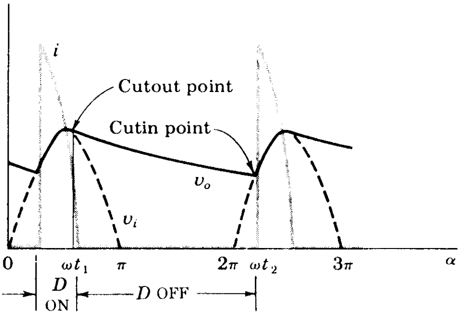


Fig. 4-25 Theoretical sketch of diode current i and output voltage v_o in a half-wave capacitor-filtered rectifier.

tion period decreases as C is made larger. It is to be emphasized that the use of a capacitor filter may impose serious restrictions on the diode, since the average current may be well within the current rating of the diode, and yet the peak current may be excessive.

Diode Nonconducting In the interval between the cutout time t_1 and the cutin time t_2, the diode is effectively out of the circuit, and the capacitor discharges through the load resistor with a time constant CR_L. Thus the capacitor voltage (equal to the load voltage) is

$$v_o = A\epsilon^{-t/CR_L} \tag{4-27}$$

To determine the value of the constant A, note from Fig. 4-25 that at the cutout time $t = t_1$, $v_o = v_i = V_m \sin \omega t_1$. Equation (4-27) thus attains the form

$$v_o = (V_m \sin \omega t_1)\epsilon^{-(t-t_1)/CR_L} \tag{4-28}$$

Since t_1 is known from Eq. (4-26), v_o can be plotted as a function of time. This exponential curve is indicated in Fig. 4-25, and where it intersects the curve $V_m \sin \omega t$ (in the following cycle) is the cutin point t_2. The validity of this statement follows from the fact that at an instant of time greater than t_2, the transformer voltage v_i (the sine curve) is greater than the capacitor voltage v_o (the exponential curve). Since the diode voltage is $v = v_i - v_o$, then v will be positive beyond t_2 and the diode will become conducting. Thus t_2 is the cutin point. No analytic expression can be given for t_2; it must be found graphically by the method outlined above.

Full-wave Circuit Consider a full-wave rectifier with a capacitor filter obtained by placing a capacitor C across R_L in Fig. 4-19. The analysis of this circuit requires a simple extension of that just made for the half-wave circuit. If in Fig. 4-25 a dashed half-sinusoid is added between π and 2π, the result is the dashed full-wave voltage in Fig. 4-26. The cutin point now lies between π and 2π, where the exponential portion of v_o intersects this sinusoid. The cutout point is the same as that found for the half-wave rectifier.

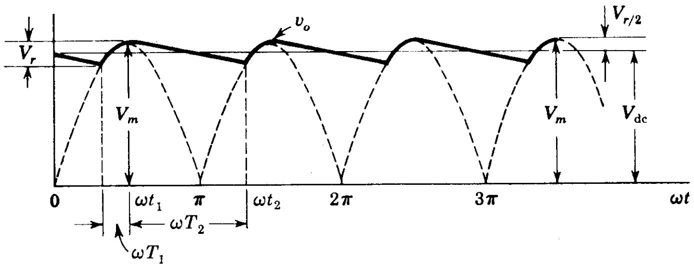


Fig. 4-26 The approximate load-voltage waveform v_o in a full-wave capacitor-filtered rectifier.

Approximate Analysis It is possible to obtain the dc output voltage for given values of the parameters ω, R_L, C, and V_m from the graphical construction indicated in Fig. 4-25. Such an analysis is involved and tedious. Hence we now present an approximate solution which is simple and yet sufficiently accurate for most engineering applications.

We assume that the output-voltage waveform of a full-wave circuit with a capacitor filter may be represented by the approximately piecewise linear curve shown in Fig. 4-26. For large values of C (so that $\omega CR_L \gg 1$) we note from Eqs. (4-25) and (4-26) that $\omega t_1 \rightarrow \pi/2$ and $v_o \rightarrow V_m$ at $t = t_1$. Also, with C very large, the exponential decay in Eq. (4-28) can be replaced by a linear fall. If the total capacitor discharge voltage (the ripple voltage) is denoted by V_r, then from Fig. 4-26, the average value of the voltage is approximately

$$V_{\mathrm{dc}} = V_m - \frac{V_r}{2} \tag{4-29}$$

It is necessary, however, to express V_r as a function of the load current and the capacitance. If T_2 represents the total nonconducting time, the capacitor, when discharging at the constant rate I_{dc}, will lose an amount of charge $I_{\mathrm{dc}}T_2$. Hence the change in capacitor voltage is $I_{\mathrm{dc}}T_2/C$, or

$$V_r = \frac{I_{\mathrm{dc}}T_2}{C} \tag{4-30}$$

The better the filtering action, the smaller will be the conduction time T_1 and the closer T_2 will approach the time of half a cycle. Hence we assume that $T_2 = T/2 = 1/2f$, where f is the fundamental power-line frequency. Then

$$V_r = \frac{I_{\mathrm{dc}}}{2fC} \tag{4-31}$$

and from Eq. (4-29),

$$V_{\mathrm{dc}} = V_m - \frac{I_{\mathrm{dc}}}{4fC} \tag{4-32}$$

This result is consistent with Thévenin's model of Fig. 4-18, with the open-circuit voltage $V = V_m$ and the effective output resistance $R_o = 1/4fC$.

The ripple is seen to vary directly with the load current I_{dc} and also inversely with the capacitance. Hence, to keep the ripple low and to ensure good regulation, very large capacitances (of the order of tens of microfarads) must be used. The most common type of capacitor for this rectifier application is the electrolytic capacitor. These capacitors are polarized, and care must be taken to insert them into the circuit with the terminal marked + to the positive side of the output.

The desirable features of rectifiers employing capacitor input filters are the small ripple and the high voltage at light load. The no-load voltage is equal, theoretically, to the maximum transformer voltage. The disadvantages of this system are the relatively poor regulation, the high ripple at large loads, and the peaked currents that the diodes must pass.

An approximate analysis similar to that given above applied to the half-wave circuit shows that the ripple, and also the drop from no load to a given load, are double the values calculated for the full-wave rectifier.

4-11 ADDITIONAL DIODE CIRCUITS

Many applications depend upon the semiconductor diode besides those already considered in this chapter. We mention four others below.

Peak Detector The half-wave capacitor-filtered rectifier circuit of Fig. 4-23 may be used to measure the peak value of an input waveform. Thus, for $R_L = \infty$, the capacitor charges to the maximum value $V_{\max}$ of v_i, the diode becomes nonconducting, and v_o remains at $V_{\max}$ (assuming an ideal capacitor with no leakage resistance shunting C). Refer to Fig. 4-24, where $V_{\max} = V_m =$ the peak value of the input sinusoid. Improved peak detector circuits are given in Sec. 16-13.

In an AM radio the amplitude of the high-frequency wave (called the *carrier*) is varied in accordance with the audio information to be transmitted. This process is called *amplitude modulation*, and such an AM waveform is illustrated in Fig. 4-27. The audio information is contained in the envelope (the locus, shown dashed, of the peak values) of the modulated waveform. The process of extracting the audio signal is called *detection*, or *demodulation*. If the input to Fig. 4-23 is the AM waveform shown in Fig. 4-27, the output v_o is the heavy-weight curve, provided that the time constant R_LC is chosen properly; that is, R_LC must be small enough so that, when the envelope decreases in magnitude, the voltage across C can fall fast enough to keep in step with the envelope, but R_LC must not be so small as to introduce excessive ripple. The order of magnitude of the frequency of an AM radio carrier is 1,000 kHz, and the audio spectrum extends from about 20 Hz to 20 kHz. Hence there should be *at least* 50 cycles of the carrier waveform for each audio cycle. If Fig. 4-27 were drawn more realistically (with a much higher ratio of carrier to audio frequency), then clearly, the ripple amplitude of the demodulated signal

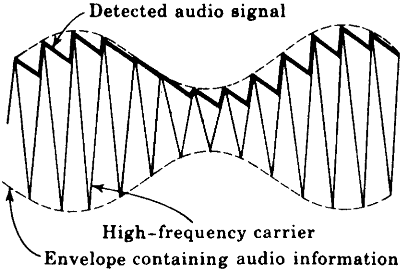


Fig. 4-27 An amplitude-modulated wave and the detected audio signal. (For ease of drawing, the carrier waveform is indicated triangular instead of sinusoidal and of much lower frequency than it really is, relative to the audio frequency.)

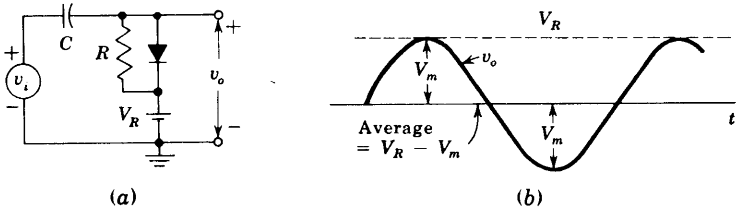


Fig. 4-28 (a) A circuit which clamps to the voltage V_R. (b) The output voltage v_o for a sinusoidal input v_i.

would be very much smaller. This low-amplitude high-frequency ripple in v_o is easily filtered so that the smoothed detected waveform is an excellent reproduction of the audio signal. The capacitor-rectifier circuit of Fig. 4-23 therefore also acts as an *envelope demodulator.*

A Clamping Circuit A function which must be frequently performed with a periodic waveform is the establishment of the recurrent positive or negative extremity at some constant reference level V_R. Such a clamping circuit is indicated in Fig. 4-28*a*. Assuming an ideal diode, the drop across the device is zero in the forward direction. Hence the output cannot rise above V_R and is said to be *clamped* to this level. If the input is sinusoidal with a peak value V_m and an average value of zero, then, as indicated in Fig. 4-28*b*, the output is sinusoidal, with an average value of $V_R - V_m$. This waveform is obtained subject to the following conditions: the diode parameters are $R_f = 0$, $R_r = \infty$, and $V_\gamma = 0$; the source impedance $R_s = 0$; and the time constant RC is much larger than the period of the signal. In practice these restrictions are not completely satisfied and the clamping is not perfect; the output voltage rises slightly above V_R, and the waveshape is a somewhat distorted version of the input.[3]

Digital-computer Circuits Since the diode is a binary device existing in either the ON or OFF state for a given interval of time, it is a very usual component in digital-computer applications. Such so-called "logic" circuits are discussed in Chap. 6 in conjunction with transistor binary applications.

Avalanche-diode Regulator The basic circuit used to decrease voltage variations across a load due to variations in output current or supply voltage is given in Fig. 3-17 and discussed in Sec. 3-11.

REFERENCES

1. Millman, J., and H. Taub: "Pulse, Digital, and Switching Waveforms," sec. 7-11, McGraw-Hill Book Company, New York, 1965.

2. Ref. 1, chap. 17.

3. Ref. 1, chap. 3.

REVIEW QUESTIONS

4-1 Explain how to obtain the dynamic characteristic from the static volt-ampere curve of a diode.

4-2 You are given the V-I output characteristic in graphical form of a new device. (*a*) Sketch the circuit using this device which will require a load-line construction to determine i and v. (*b*) Is the load line vertical, horizontal, at 135° or 45° for infinite load resistance? (*c*) For zero load resistance?

4-3 (*a*) Draw the piecewise linear volt-ampere characteristic of a *p-n* diode. (*b*) What is the circuit model for the ON state? (*c*) The OFF state?

4-4 Consider a circuit consisting of a diode D, a resistance R, and a signal source v_i in series. Define (*a*) static characteristic; (*b*) dynamic characteristic; (*c*) *transfer*, or *transmission*, characteristic. (*d*) What is the correlation between (*b*) and (*c*)?

4-5 Consider the circuit of Rev. 4-4 using a silicon diode and a 100-Ω resistance. (*a*) Plot approximately the dynamic curve. (*b*) If the input is $v_i = -1 + A \sin \omega t$, plot the current waveform for $A = 1.5$, 2.0, and 5.0 V.

4-6 For the circuit of Rev. 4-4, $v_i = V_m \sin \omega t$. If the diode is represented by its piecewise linear model, find the current as a function of time and plot.

4-7 What is meant by *break region* of a diode?

4-8 In analyzing a circuit containing several diodes by the piecewise linear method, you assume (guess) that certain of the diodes are ON and others are OFF. Explain carefully how you determine whether or not the assumed state of each diode is correct.

4-9 Consider a series circuit consisting of a diode D, a resistance R, a reference battery V_R, and an input signal v_i. The output is taken across R and V_R in series. Draw the transfer characteristic if the anode of D is connected to the positive terminal of the battery. Use the piecewise linear diode model.

4-10 Repeat Rev. 4-9 with the anode of D connected to the negative terminal of the battery.

4-11 If v_i is sinusoidal and D is ideal (with $R_f = 0$, $V_\gamma = 0$, and $R_r = \infty$), find the output waveforms in (*a*) Rev. 4-9 and (*b*) Rev. 4-10.

4-12 Sketch the circuit of a *double-ended clipper* using ideal *p-n* diodes which limit the output between ±10 V.

4-13 Repeat Rev. 4-12 using avalanche diodes.

4-14 (*a*) What is a *comparator circuit?* (*b*) How does such a circuit differ from a *clipping circuit?*

4-15 (*a*) What is a *sampling gate?* (*b*) Sketch the circuit of a four-diode sampling gate.

4-16 Define in words and as an equation (*a*) dc current I_{dc}; (*b*) dc voltage V_{dc}; (*c*) ac current I_{rms}.

4-17 (*a*) Sketch the circuit for a full-wave rectifier. (*b*) Derive the expression for (1) the dc current; (2) the dc load voltage; (3) the dc diode voltage; (4) the rms current.

4-18 (*a*) Define *regulation.* (*b*) Derive the regulation equation for a full-wave circuit.

4-19 Draw the Thévenin's model for a full-wave rectifier.

4-20 (*a*) Define *peak inverse voltage.* (*b*) What is the peak inverse voltage for a full-wave circuit using ideal diodes? (*c*) Repeat part *b* for a half-wave rectifier.

4-21 Sketch the circuit of a bridge rectifier and explain its operation.

4-22 Repeat Rev. 4-21 for a rectifier meter circuit.

4-23 Repeat Rev. 4-21 for a voltage multiplier.

4-24 (*a*) Draw the circuit of a half-wave capacitive rectifier. (*b*) At no load draw the steady-state voltage across the capacitor and also across the diode.

4-25 (*a*) Draw the circuit of a full-wave capacitive rectifier. (*b*) Draw the output voltage under load. Indicate over what period of time the diode conducts. Make no calculations. (*c*) Indicate the diode current waveform superimposed upon the output waveform.

4-26 For the circuit of Rev. 4-25, derive the expression for (*a*) the diode current; (*b*) the cutout angle. (*c*) How is the cutin angle found?

4-27 (*a*) Consider a full-wave capacitor rectifier using a large capacitance C. Sketch the approximate output waveform. (*b*) Derive the expression for the peak ripple voltage. (*c*) Derive the Thévenin's model for this rectifier.

4-28 For a full-wave capacitor rectifier circuit, list (*a*) two advantages; (*b*) three disadvantages.

4-29 Describe (*a*) *amplitude modulation* and (*b*) *detection.*

5 / TRANSISTOR CHARACTERISTICS

The physical behavior of a semiconductor triode, called a *bipolar junction transistor* (BJT), is given. The volt-ampere characteristics of this device are studied. It is demonstrated that the transistor is capable of producing amplification. Analytical expressions relating the transistor currents with the junction voltages are indicated. Typical voltage values are given, for the several possible modes of operation.

5-1 THE JUNCTION TRANSISTOR[1]

A junction transistor consists of a silicon (or germanium) crystal in which a layer of *n*-type silicon is sandwiched between two layers of *p*-type silicon. Alternatively, a transistor may consist of a layer of *p*-type between two layers of *n*-type material. In the former case the transistor is referred to as a *p-n-p* transistor, and in the latter case, as an *n-p-n* transistor. The semiconductor sandwich is extremely small, and is hermetically sealed against moisture inside a metal or plastic case. Manufacturing techniques and constructional details for several transistor types are described in Sec. 5-4.

The two types of transistor are represented in Fig. 5-1*a*. The representations employed when transistors are used as circuit elements are shown in Fig. 5-1*b*. The three portions of a transistor are known as *emitter*, *base*, and *collector*. The arrow on the emitter lead specifies the direction of current flow when the emitter-base junction is biased in the forward direction. In *both* cases, however, the emitter, base, and collector currents, I_E, I_B, and I_C, respectively, are assumed positive when the currents flow *into* the transistor. The symbols V_{EB}, V_{CB}, and V_{CE} are the emitter-base, collector-base, and collector-emitter voltages, respectively. (More specifically, V_{EB} represents the voltage *drop* from emitter to base.)

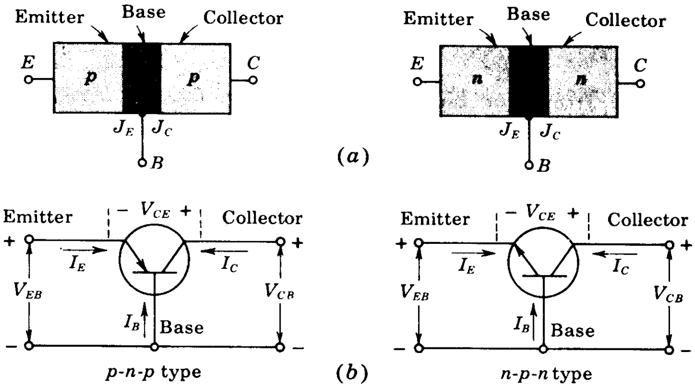

Emitter
Base
Collector
E
p
n
p
C
J_E
J_C
B
(a)
Emitter
Base
Collector
E
n
p
n
C
J_E
J_C
B
Emitter
$- V_{CE} +$
Collector
+
I_E
I_C
+
V_{EB}
V_{CB}
I_B
Base
−
−
p-n-p type
(b)
n-p-n type

Fig. 5-1 (a) A *p-n-p* and an *n-p-n* transistor. The emitter (collector) junction is $J_E(J_C)$. (b) Circuit representation of the two transistor types.

Open-circuited Transistor If no external biasing voltages are applied, all transistor currents must be zero. The potential barriers at the junctions adjust to the contact difference of potential V_o—given in Eq. (2-63) (a few tenths of a volt)—required so that no free carriers cross each junction. If we assume a completely symmetrical junction (emitter and collector regions having identical physical dimensions and doping concentrations), the barrier height is identical at the emitter junction J_E and at the collector junction J_C, as indicated in Fig. 5-2*a*. The narrow space-charge regions at the junctions have been neglected.

Under open-circuited conditions, the minority concentration is constant within each section and is equal to its thermal-equilibrium value, n_{po} in the *p*-type emitter and collector regions and p_{no} in the *n*-type base, as shown in Fig. 5-2*b*. Since the transistor may be looked upon as a *p-n* diode followed by an *n-p* diode, much of the theory developed in the preceding chapters for the junction diode will be used to explain the physical behavior of the transistor,

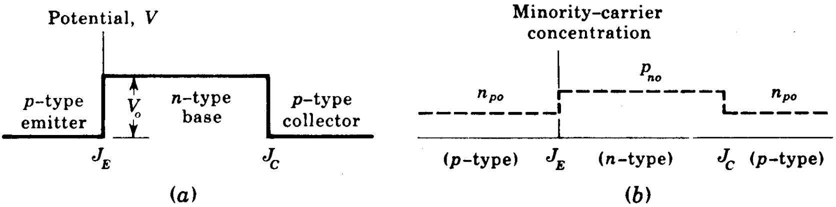

Potential, V
p-type emitter
V_o
n-type base
p-type collector
J_E
J_C
(a)
Minority-carrier concentration
p_{no}
n_{po}
n_{po}
(p-type)
J_E
(n-type)
J_C
(p-type)
(b)

Fig. 5-2 (a) The potential and (b) the minority-carrier density in each section of an open-circuited symmetrical *p-n-p* transistor.

when voltages are applied so as to disturb it from the equilibrium situation pictured in Fig. 5-2.

The Transistor Biased in the Active Region We may now begin to appreciate the essential features of a transistor as an active circuit element by considering the situation depicted in Fig. 5-3*a*. Here a *p*-*n*-*p* transistor is shown with voltage sources which serve to bias the emitter-base junction in the forward direction and the collector-base junction in the reverse direction. The potential variation through the biased transistor is indicated in Fig. 5-3*b*. The dashed curve applies to the case before the application of external biasing voltages (Fig. 5-2*a*), and the solid curve to the case after the biasing voltages are applied. The externally applied voltages appear, essentially, across the junctions. Hence, as shown in Fig. 5-3*b*, the forward biasing of the emitter junction lowers the emitter-base potential barrier by $|V_{EB}|$, whereas the reverse biasing of the collector junction increases the collector-base potential barrier by $|V_{CB}|$. The lowering of the emitter-base barrier permits minority-carrier injection; that is, holes are injected into the base, and electrons are injected into the emitter region. The excess holes diffuse across the *n*-type base,

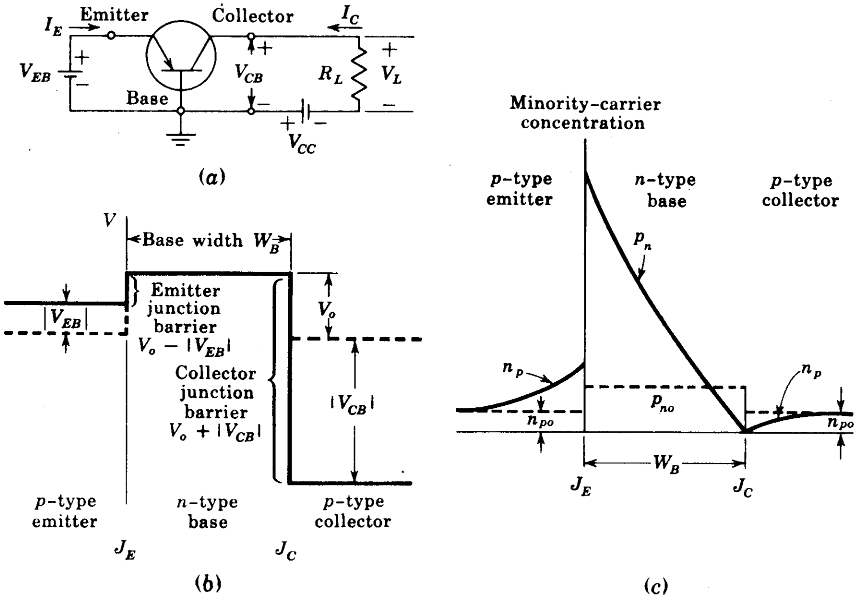

I_E
Emitter
Collector
I_C
V_{EB}
+
−
Base
V_{CB}
R_L
V_L
V_{CC}
(a)
V
Base width W_B
$|V_{EB}|$
Emitter junction barrier $V_o - |V_{EB}|$
V_o
Collector junction barrier $V_o + |V_{CB}|$
$|V_{CB}|$
p-type emitter
n-type base
p-type collector
J_E
J_C
(b)
Minority-carrier concentration
p-type emitter
n-type base
p-type collector
p_n
n_p
p_{no}
n_{po}
n_p
n_{po}
W_B
J_E
J_C
(c)

Fig. 5-3 (*a*) A *p*-*n*-*p* transistor biased in the active region (the emitter is forward-biased and the collector is reverse-biased). (*b*) The potential variation through the transistor. The narrow depletion regions at the junctions are negligibly small. (*c*) The minority-carrier concentration in each section of the transistor. It is assumed that the emitter is much more heavily doped than the base.

where the electric field intensity $\mathcal{E}$ is zero, to the collector junction. At J_C the field is positive and large ($\mathcal{E} = -dV/dx \gg 0$), and hence holes are accelerated across this junction. In other words, the holes which reach J_C fall down the potential barrier, and are therefore *collected* by the collector. Since the applied potential across J_C is negative, then from the law of the junction, Eq. (3-4), p_n is reduced to zero at the collector as shown in Fig. 5-3*c*. Similarly, the reverse collector-junction bias reduces the collector electron density n_p to zero at J_C. The minority-carrier-density curves pictured in Fig. 5-3*c* should be compared with the corresponding concentration plots for the forward- and reverse-biased *p*-*n* junction given in Fig. 3-14.

5-2 TRANSISTOR CURRENT COMPONENTS

In Fig. 5-4 we show the various current components which flow across the forward-biased emitter junction and the reverse-biased collector junction. The emitter current I_E consists of hole current I_{pE} (holes crossing from emitter into base) and electron current I_{nE} (electrons crossing from base into the emitter). The ratio of hole to electron currents, I_{pE}/I_{nE}, crossing the emitter junction is proportional to the ratio of the conductivity of the *p* material to that of the *n* material. In a commercial transistor the doping of the emitter is made much larger than the doping of the base. This feature ensures (in a *p*-*n*-*p* transistor) that the emitter current consists almost entirely of holes.

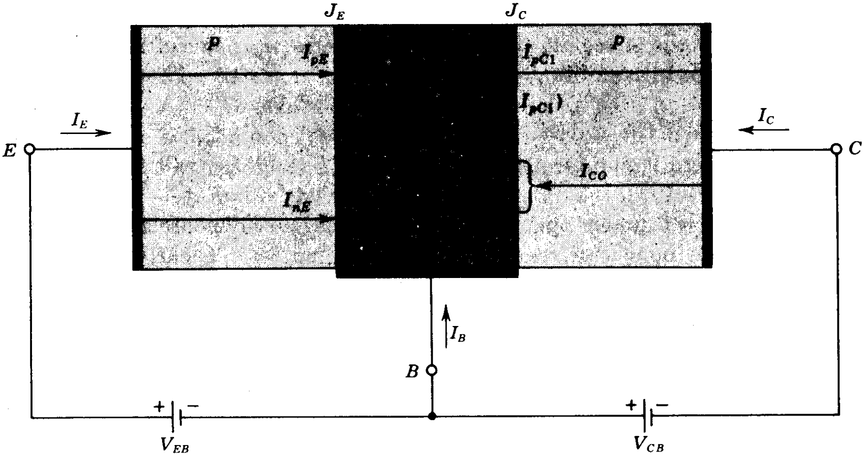


Fig. 5-4 Transistor current components for a forward-biased emitter junction and a reversed-biased collector junction. If a current has a subscript $p(n)$, it consists of holes (electrons) moving in the same (opposite) direction as the arrow indicating the current direction.

Such a situation is desirable since the current which results from electrons crossing the emitter junction from base to emitter does not contribute carriers which can reach the collector.

We assume low-level injection, and hence (Sec. 2-11) the minority current I_{pE} is the hole *diffusion* current into base and its magnitude is proportional to the slope at J_E of the p_n curve [Eq. (2-36)]. Similarly, I_{nE} is the electron *diffusion* current into the emitter, and its magnitude is proportional to the slope at J_E of the n_p curve. Note that I_{pE} and I_{nE} correspond to the minority-carrier diffusion currents $I_{pn}(0)$ and $I_{np}(0)$, respectively, in Fig. 3-4 for the currents crossing a *p-n* junction. Just as the total current crossing the junction in Fig. 3-4 is $I = I_{pn}(0) + I_{np}(0)$, so the total emitter current in Fig. 5-4 is

$$I_E = I_{pE} + I_{nE} \tag{5-1}$$

All currents in this equation are positive for a *p-n-p* transistor.

Not all the holes crossing the emitter junction J_E reach the collector junction J_C, because some of them combine with the electrons in the *n*-type base. In Fig. 5-4, let I_{pC1} represent the hole current at J_C as a result of holes crossing the base from the emitter. Hence there must be a bulk recombination hole current $I_{pE} - I_{pC1}$ leaving the base, as indicated in Fig. 5-4 (actually, electrons enter the base region from the external circuit through the base lead to supply those charges which have been lost by recombination with the holes injected into the base across J_E).

Consider, for the moment, an open-circuited emitter, while the collector junction remains reverse-biased. Then I_C must equal the reverse saturation current I_{CO} of the back-biased diode at J_C. This *reverse* current consists of two components, as shown in Fig. 5-4, I_{nCO} consisting of electrons moving from the *p* to the *n* region across J_C and a term, I_{pCO}, resulting from holes crossing from *n* to *p* across J_C.

$$-I_{CO} = I_{nCO} + I_{pCO} \tag{5-2}$$

(The minus sign is chosen arbitrarily so that I_C and I_{CO} will have the same sign.)

Since $I_E = 0$ under open-circuit conditions, no holes are injected across J_E, and hence none can reach J_C from the emitter. Clearly, I_{pCO} results from the small concentration of holes generated thermally within the base.

Now let us return to the situation depicted in Fig. 5-4, where the emitter is forward-biased so that

$$I_C = I_{CO} - I_{pC1} = I_{CO} - \alpha I_E \tag{5-3}$$

where α is defined as the fraction of the total emitter current [given in Eq. (5-1)] which represents holes which have traveled from the emitter across the base to the collector. For a *p-n-p* transistor, I_E is positive and both I_C and I_{CO} are negative, which means that the current in the collector lead is in the direction opposite to that indicated by the arrow of I_C in Fig. 5-4. For an *n-p-n* transistor these currents are reversed.

The electron current crossing J_C is I_{nCO} and represents electrons diffusing from the collector into the base (and hence a positive current from the base into the collector), and its magnitude is proportional to the slope at J_C of the n_p distribution in Fig. 5-3*c*. The total diffusion hole current crossing J_C from the base is

$$I_{pC} \equiv I_{pC1} + I_{pCO} \tag{5-4}$$

and its magnitude is proportional to the slope at J_C of the p_n distribution in Fig. 5-3*c*.

Large-signal Current Gain α From Eq. (5-3) it follows that α may be defined as the ratio of the negative of the collector-current increment from cutoff ($I_C = I_{CO}$) to the emitter-current change from cutoff ($I_E = 0$), or

$$\alpha \equiv -\frac{I_C - I_{CO}}{I_E - 0} \tag{5-5}$$

Alpha is called the *large-signal current gain* of a common-base transistor. Since I_C and I_E have opposite signs (for either a *p-n-p* or an *n-p-n* transistor), then α, as defined, is always positive. Typical numerical values of α lie in the range 0.90 to 0.995. It should be pointed out that α is not a constant, but varies with emitter current I_E, collector voltage V_{CB}, and temperature.

A Generalized Transistor Equation Equation (5-3) is valid only in the *active region*, that is, if the emitter is forward-biased and the collector is reverse-biased. For this mode of operation the collector current is essentially independent of collector voltage and depends only upon the emitter current. Suppose now that we seek to generalize Eq. (5-3) so that it may apply not only when the collector junction is substantially reverse-biased, but also for any voltage across J_C. To achieve this generalization we need only replace I_{CO} by the current in a *p-n* diode (that consisting of the base and collector regions). This current is given by the volt-ampere relationship of Eq. (3-7), with I_o replaced by $-I_{CO}$ and V by V_C, where the symbol V_C represents the drop across J_C from the p to the n side. The complete expression for I_C for any V_C and I_E is

$$I_C = -\alpha I_E + I_{CO}(1 - \epsilon^{V_C/V_T}) \tag{5-6}$$

Note that if V_C is negative and has a magnitude large compared with V_T, Eq. (5-6) reduces to Eq. (5-3). The physical interpretation of Eq. (5-6) is that the *p-n* junction diode current crossing the collector junction is augmented by the fraction α of the current I_E flowing in the emitter. This relationship is derived in Sec. 19-14 by evaluating quantitatively the various current components introduced above.

The similarity between Eq. (5-6) and Eq. (3-34) for the photodiode should be noted. Both equations represent the volt-ampere characteristics of a reverse-biased diode (with reverse saturation current $I_O = I_{CO}$) subjected to external excitation. In the case of the photodiode, the external stimulus is radiation, which injects minority carriers into the semiconductor, resulting in the current I_s. This component of current, which is proportional to the light intensity, augments the current from the simple *p-n* junction diode. Similarly, for the transistor, the collector-base diode current is increased by the term $-\alpha I_E$, which is proportional to the "external excitation," namely, the emitter current resulting from the voltage applied between emitter and base.

5-3 THE TRANSISTOR AS AN AMPLIFIER

A load resistor R_L is in series with the collector supply voltage V_{CC} of Fig. 5-3*a*. A small voltage change ΔV_i between emitter and base causes a relatively large emitter-current change ΔI_E. We define by the symbol α' that fraction of this current change which is collected and passes through R_L, or $\Delta I_C = \alpha' \Delta I_E$. The change in output voltage across the load resistor

$$\Delta V_L = -R_L \Delta I_C = -\alpha' R_L \Delta I_E$$

may be many times the change in input voltage ΔV_i. Under these circumstances, the voltage amplification $A \equiv \Delta V_L / \Delta V_i$ will be greater than unity, and the transistor acts as an amplifier. If the dynamic resistance of the emitter junction is r_e, then $\Delta V_i = r_e \Delta I_E$, and

$$A \equiv -\frac{\alpha' R_L \Delta I_E}{r_e \Delta I_E} = -\frac{\alpha' R_L}{r_e} \tag{5-7}$$

From Eq. (3-14), $r_e = 26/I_E$, where I_E is the quiescent emitter current in milliamperes. For example, if $r_e = 40\ \Omega$, $\alpha' = -1$, and $R_L = 3{,}000\ \Omega$, $A = +75$. This calculation is oversimplified, but in essence it is correct and gives a physical explanation of why the transistor acts as an amplifier. The transistor provides power gain as well as voltage or current amplification. From the foregoing explanation it is clear that current in the low-resistance input circuit is transferred to the high-resistance output circuit. The word "transistor," which originated as a contraction of "transfer resistor," is based upon the above physical picture of the device.

The Parameter α' The parameter α' introduced above is defined as the ratio of the change in the collector current to the change in the emitter current at constant collector-to-base voltage and is called the *negative of the small-signal short-circuit current transfer ratio, or gain*. More specifically,

$$\alpha' \equiv \left.\frac{\Delta I_C}{\Delta I_E}\right|_{V_{CB}} \tag{5-8}$$

On the assumption that α is independent of I_E, then from Eq. (5-3) it follows that $\alpha' = -\alpha$.

5-4 TRANSISTOR CONSTRUCTION

Four basic techniques have been developed for the manufacture of diodes, transistors, and other semiconductor devices. Consequently, such devices may be classified[2,3] into one of the following types: grown, alloy, diffusion, or epitaxial.

Grown Type The *n-p-n* grown-junction transistor is illustrated in Fig. 5-5*a*. It is made by drawing a single crystal from a melt of silicon or germanium whose impurity concentration is changed during the crystal-drawing operation by adding *n*- or *p*-type atoms as required.

Alloy Type This technique, also called the *fused* construction, is illustrated in Fig. 5-5 for a *p-n-p* transistor. The center (base) section is a thin wafer of *n*-type material. Two small dots of indium are attached to opposite sides of the wafer, and the whole structure is raised for a short time to a high temperature, above the melting point of indium but below that of germanium. The indium dissolves the germanium beneath it and forms a saturation solution. On cooling, the germanium in contact with the base material recrystallizes, with enough indium concentration to change it from *n* to *p* type. The collector is made larger than the emitter, so that the collector subtends a large angle as viewed from the emitter. Because of this geometrical arrangement, very little emitter current follows a diffusion path which carries it to the base rather than to the collector.

Diffusion Type This technique consists of subjecting a semiconductor wafer to gaseous diffusions of both *n*- and *p*-type impurities to form both the emitter and the collector junctions. A *planar* silicon transistor of the diffusion

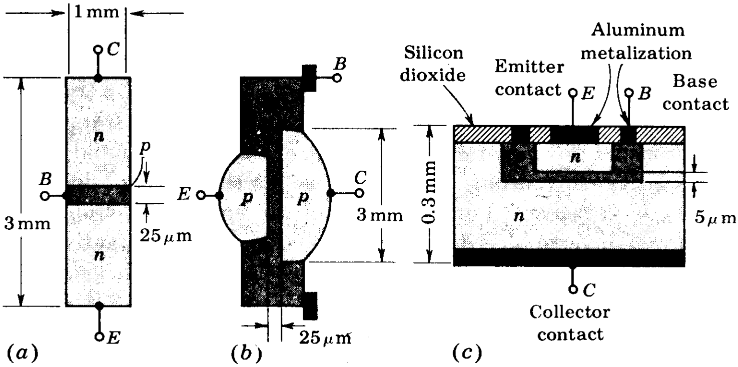


Fig. 5-5 Construction of transistors. (a) Grown, (b) alloy, and (c) diffused planar types. (The dimensions are approximate, and the figures are not drawn to scale.)

type is illustrated in Fig. 5-5*c*. In this process (described in greater detail in Chap. 7 on integrated-circuit fabrication), the base-collector junction area is determined by a diffusion mask. The emitter is then diffused on the base through a different mask. A thin layer of silicon dioxide is grown over the entire surface and photoetched, so that aluminum contacts can be made for the emitter and base leads (Fig. 5-5*c*). Because of the passivating action of this oxide layer, most surface problems are avoided and very low leakage currents result. There is also an improvement in the current gain at low currents and in the noise figure.

Epitaxial Type The epitaxial technique (Sec. 7-2) consists of growing a very thin, high-purity, single-crystal layer of silicon or germanium on a heavily doped substrate of the same material. This augmented crystal forms the collector on which the base and emitter may be diffused (Fig. 7-6).

The foregoing techniques may be combined to form a large number of methods for constructing transistors. For example, there are *diffused-alloy* types, *grown-diffused* devices, *alloy-emitter–epitaxial-base* transistors, etc. The special features of transistors of importance at high frequencies are discussed in Chap. 11. The volt-ampere characteristics at low frequencies of all types of junction transistors are essentially the same, and the discussion to follow applies to them all.

Finally, because of its historical significance, let us mention the first type of transistor to be invented. This device consists of two sharply pointed tungsten wires pressed against a semiconductor wafer. However, the reliability and reproducibility of such point-contact transistors are very poor, and as a result these transistors are no longer of practical importance.

5-5 THE COMMON-BASE CONFIGURATION

In Fig. 5-3*a*, a *p-n-p* transistor is shown in a *grounded-base* configuration. This circuit is also referred to as a *common-base*, or CB, configuration, since the base is common to the input and output circuits. For a *p-n-p* transistor the largest current components are due to holes. Since holes flow from the emitter to the collector and down toward ground out of the base terminal, then, referring to the polarity conventions of Fig. 5-1, we see that I_E is positive, I_C is negative, and I_B is negative. For a forward-biased emitter junction, V_{EB} is positive, and for a reverse-biased collector junction, V_{CB} is negative. For an *n-p-n* transistor all current and voltage polarities are the negative of those for a *p-n-p* transistor.

From Eq. (5-6) we see that the output (collector) current I_C is completely determined by the input (emitter) current I_E and the output (collector-to-base) voltage $V_{CB} = V_C$. This output relationship may be written in implicit form as

$$I_C = \phi_2(V_{CB}, I_E) \tag{5-9}$$

(This equation is read "I_C is some function ϕ_2 of V_{CB} and I_E.")

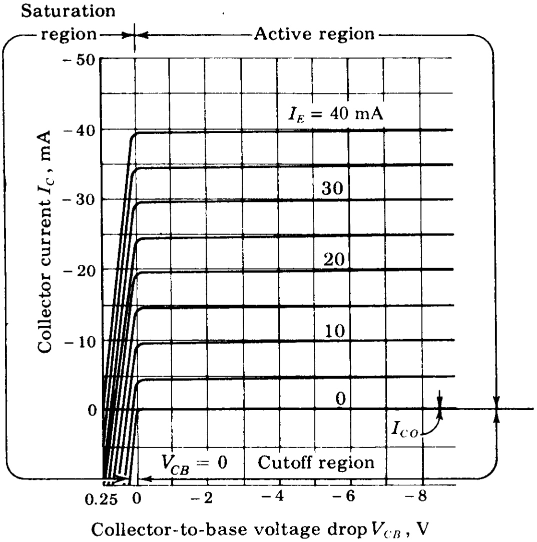


Fig. 5-6 Typical common-base output characteristics of a *p-n-p* transistor. The cutoff, active, and saturation regions are indicated. Note the expanded voltage scale in the saturation region.

Similarly, if we continue to choose V_{CB} and I_E as independent variables, the input (emitter-to-base) voltage V_{EB} is completely determined from these two variables. In implicit form the input characteristic is given by

$$V_{EB} = \phi_1(V_{CB}, I_E) \tag{5-10}$$

The relation of Eq. (5-9) is given in Fig. 5-6 for a typical *p-n-p* germanium transistor and is a plot of collector current I_C versus collector-to-base voltage drop V_{CB}, with emitter current I_E as a parameter. The curves of Fig. 5-6 are known as the *output*, or *collector, static characteristics*. The relation of Eq. (5-10) is given in Fig. 5-7 for the same transistor, and is a plot of emitter-to-base voltage V_{EB} versus emitter current I_E, with collector-to-base voltage V_{CB} as a parameter. This set of curves is referred to as the *input*, or *emitter, static characteristics*. We digress now to discuss a phenomenon which is used to account for the shapes of the transistor characteristics.

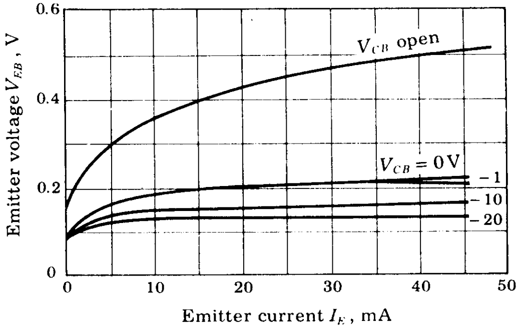


Fig. 5-7 Common-base input characteristics of a typical *p-n-p* germanium junction transistor.

The Early Effect, or Base-width Modulation[4] In Fig. 5-3 the narrow space-charge regions in the neighborhood of the junctions are neglected. This restriction is now to be removed. From Eq. (3-21) we note that the width W of the depletion region of a diode increases with the magnitude of the reverse voltage. Since the emitter junction is forward-biased but the collector junction is reverse-biased in the active region, then in Fig. 5-8 the barrier width at J_E is negligible compared with the space-charge with W at J_C.

The transition region at a junction is the region of uncovered charges on both sides of the junction at the positions occupied by the impurity atoms. As the voltage applied across the junction increases, the transition region penetrates deeper into the collector and base. Because neutrality of charge must be maintained, the number of uncovered charges on each side remains equal. Since the doping in the base is ordinarily substantially smaller than that of the collector, the penetration of the transition region into the base is much larger than into the collector. Hence the collector depletion region is neglected in Fig. 5-8, and all the immobile charge is indicated in the base region.

If the metallurgical base width is W_B, then the effective electrical base width is $W'_B = W_B - W$. This modulation of the effective base width by the collector voltage is known as the *Early effect*. The decrease in W'_B with increasing reverse collector voltage has three consequences: First, there is less chance for recombination within the base region. Hence α increases with increasing $|V_{CB}|$. Second, the concentration gradient of minority carriers p_n is increased within the base, as indicated in Fig. 5-8b. Note that p_n becomes

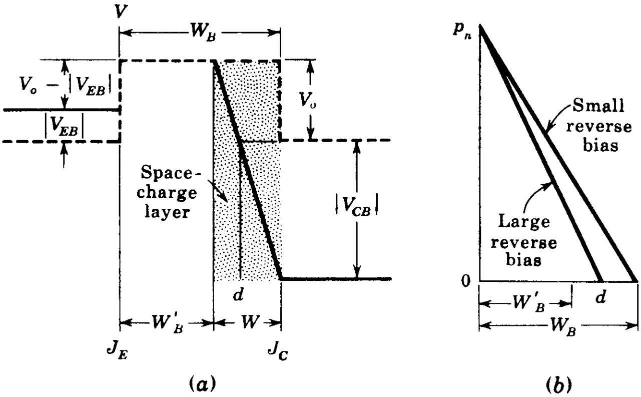


Fig. 5-8 (a) The potential variation through a p-n-p transistor. The space-charge width W at the collector junction increases, and hence the effective base width W'_B decreases with increasing $|V_{CB}|$. (Compare with Fig. 5-3.) (b) The injected minority-carrier charge density within the base.

zero at the distance d (between W'_B and W_B), where the potential with respect to the base falls below V_o. At this distance the effective applied potential becomes negative, and the law of the junction, Eq. (3-4), yields $p_n = 0$. Since the hole current injected across the emitter is proportional to the gradient of p_n at J_E, then I_E increases with increasing reverse collector voltage. Third, for extremely large voltages, W'_B may be reduced to zero, as in Fig. 5-24, causing voltage breakdown in the transistor. This phenomenon of *punch-through* is discussed further in Sec. 5-13.

The Input Characteristics A qualitative understanding of the form of the input and output characteristics is not difficult if we consider the fact that the transistor consists of two diodes placed in series "back to back" (with the two cathodes connected together). In the active region the input diode (emitter-to-base) is biased in the forward direction. The input characteristics of Fig. 5-7 represent simply the forward characteristic of the emitter-to-base diode for various collector voltages. A noteworthy feature of the input characteristics is that there exists a *cutin*, *offset*, or *threshold*, voltage V_γ below which the emitter current is very small. In general, V_γ is approximately 0.1 V for germanium transistors (Fig. 5-7) and 0.5 V for silicon.

The shape of the input characteristics can be understood if we consider the fact that an increase in magnitude of collector voltage will, by the Early effect, cause the emitter current to increase, with V_{EB} held constant. Thus the curves shift downward as $|V_{CB}|$ increases, as noted in Fig. 5-7. The curve with the collector open represents the characteristic of the forward-biased emitter diode.

The Output Characteristics Note, as in Fig. 5-6, that it is customary to plot along the abscissa and to the right that polarity of V_{CB} which reverse-biases the collector junction even if this polarity is negative. If $I_E = 0$, the collector current is $I_C = I_{CO}$. For other values of I_E, the output-diode reverse current is augmented by the fraction of the input-diode forward current which reaches the collector. Note also that I_C and I_{CO} are negative for a *p-n-p* transistor and positive for an *n-p-n* transistor.

Active Region In this region *the collector junction is biased in the reverse direction and the emitter junction in the forward direction.* Consider first that the emitter current is zero. Then the collector current is small and equals the reverse saturation current I_{CO} (microamperes for germanium and nanoamperes for silicon) of the collector junction considered as a diode. Suppose now that a forward emitter current I_E is caused to flow in the emitter circuit. Then a fraction $-\alpha I_E$ of this current will reach the collector, and I_C is therefore given by Eq. (5-3). In the active region, the collector current is essentially independent of collector voltage and depends only upon the emitter current. However, because of the Early effect, we note in Fig. 5-6 that there actually is a small (perhaps 0.5 percent) increase in $|I_C|$ with $|V_{CB}|$. Because α is less than, but almost equal to, unity, the magnitude of the collector current is (slightly) less than that of the emitter current.

Saturation Region The region to the left of the ordinate, $V_{CB} = 0$, and above the $I_E = 0$ characteristics, in which *both emitter and collector junctions are forward-biased*, is called the *saturation* region. We say that "bottoming" has taken place because the voltage has fallen near the bottom of the characteristic where $V_{CB} \approx 0$. Actually, V_{CB} is slightly positive (for a *p-n-p* transistor) in this region, and this forward biasing of the collector accounts for the large change in collector current with small changes in collector voltage. For a forward bias, I_C increases exponentially with voltage according to the diode relationship [Eq. (3-9)]. A forward bias means that the collector *p* material is made positive with respect to the base *n* side, and hence that hole current flows from the *p* side across the collector junction to the *n* material. This hole flow corresponds to a positive change in collector current. Hence the collector current increases rapidly, and as indicated in Fig. 5-6, I_C may even become positive if the forward bias is sufficiently large.

Cutoff Region The characteristic for $I_E = 0$ passes through the origin, but is otherwise similar to the other characteristics. This characteristic is not coincident with the voltage axis, though the separation is difficult to show because I_{CO} is only a few nanoamperes or microamperes. The region below the $I_E = 0$ characteristic, for which the *emitter and collector junctions are both reverse-biased*, is referred to as the *cutoff* region. The temperature characteristics of I_{CO} are discussed in Sec. 5-7.

5-6 THE COMMON-EMITTER CONFIGURATION

Most transistor circuits have the emitter, rather than the base, as the terminal common to both input and output. Such a *common-emitter* (CE), or *grounded-emitter*, configuration is indicated in Fig. 5-9. In the common-emitter (as in the common-base) configuration, the input current and the output voltage are taken as the independent variables, whereas the input voltage and output

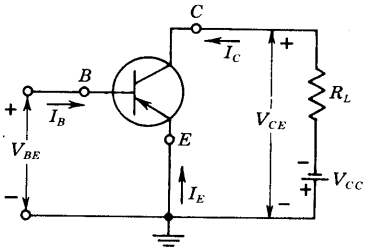


Fig. 5-9 A transistor common-emitter configuration. The symbol V_{CC} is a positive number representing the magnitude of the supply voltage.

current are the dependent variables. We may write

$$V_{BE} = f_1(V_{CE}, I_B) \tag{5-11}$$

$$I_C = f_2(V_{CE}, I_B) \tag{5-12}$$

Equation (5-11) describes the family of input characteristic curves, and Eq. (5-12) describes the family of output characteristic curves. Typical output and input characteristic curves for a *p-n-p* junction germanium transistor are given in Figs. 5-10 and 5-11, respectively. In Fig. 5-10 the abscissa is the collector-to-emitter voltage V_{CE}, the ordinate is the collector current I_C, and the curves are given for various values of base current I_B. For a fixed value of I_B, the collector current is not a very sensitive value of V_{CE}. However, the slopes of the curves of Fig. 5-10 are larger than in the common-base characteristics of Fig. 5-6. Observe also that the base current is much smaller than the emitter current.

The locus of all points at which the collector dissipation is 150 mW is indicated in Fig. 5-10 by a solid line $P_C = 150$ mW. This curve is the hyperbola $P_C = V_{CB}I_C \approx V_{CE}I_C =$ constant. To the right of this curve the rated collector dissipation is exceeded. In Fig. 5-10 we have selected $R_L = 500\ \Omega$ and a supply $V_{CC} = 10$ V and have superimposed the corresponding load line on the output characteristics. The method of constructing a load line is identical with that explained in Sec. 4-2 in connection with a diode.

The Input Characteristics In Fig. 5-11 the abscissa is the base current I_B, the ordinate is the base-to-emitter voltage V_{BE}, and the curves are given for various values of collector-to-emitter voltage V_{CE}. We observe that, with the collector shorted to the emitter and the emitter forward-biased, the input characteristic is essentially that of a forward-biased diode. If V_{BE} becomes zero, then I_B will be zero, since under these conditions both emitter and collector junctions will be short-circuited. In general, increasing $|V_{CE}|$ with constant

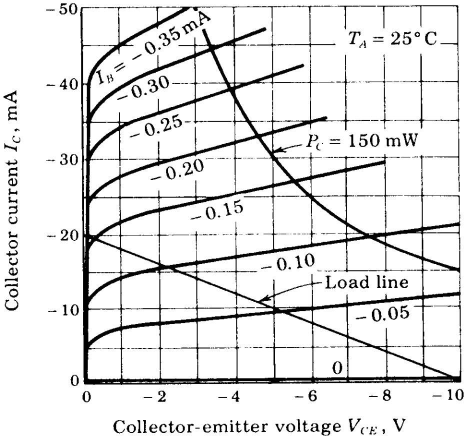


Fig. 5-10 Typical common-emitter output characteristics of a *p-n-p* germanium junction transistor. A load line corresponding to $V_{CC} = 10$ V and $R_L = 500\ \Omega$ is superimposed. (Courtesy of Texas Instruments, Inc.)

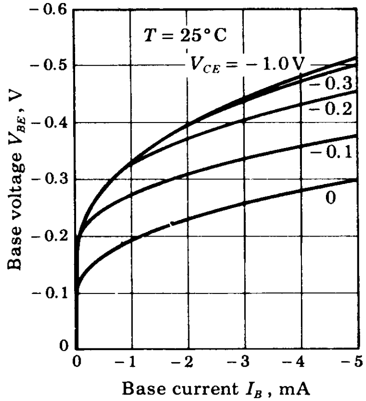


Fig. 5-11 Typical common-emitter input characteristics of the *p-n-p* germanium junction transistor of Fig. 5-10.

V_{BE} causes a decrease in base width W'_B (Fig. 5-8) and results in a decreasing recombination base current. These considerations account for the shape of input characteristics shown in Fig. 5-11.

The input characteristics for silicon transistors are similar in form to those in Fig. 5-11. The only notable difference in the case of silicon is that the curves break away from zero current in the range 0.5 to 0.6 V, rather than in the range 0.1 to 0.2 V as for germanium.

The Output Characteristics This family of curves may be divided into three regions, just as was done for the CB configuration. The first of these, the *active region*, is discussed here, and the *cutoff* and *saturation regions* are considered in the next two sections.

In the active region *the collector junction is reverse-biased and the emitter junction is forward-biased.* In Fig. 5-10 the active region is the area to the right of the ordinate V_{CE} = a few tenths of a volt and above $I_B = 0$. In this region the transistor output current responds most sensitively to an input signal. If the transistor is to be used as an amplifying device without appreciable distortion, it must be restricted to operate in this region.

The common-emitter characteristics in the active region are readily understood qualitatively on the basis of our earlier discussion of the common-base configuration. From Kirchhoff's current law (KCL) applied to Fig. 5-9, the base current is

$$I_B = -(I_C + I_E) \tag{5-13}$$

Combining this equation with Eq. (5-3), we find

$$I_C = \frac{I_{CO}}{1-\alpha} + \frac{\alpha I_B}{1-\alpha} \tag{5-14}$$

If we define β by

$$\beta \equiv \frac{\alpha}{1-\alpha} \tag{5-15}$$

then Eq. (5-14) becomes

$$I_C = (1+\beta)I_{CO} + \beta I_B \tag{5-16}$$

Note that usually $I_B \gg I_{CO}$, and hence $I_C \approx \beta I_B$ in the active region.

If α were truly constant, then, according to Eq. (5-14), I_C would be independent of V_{CE} and the curves of Fig. 5-10 would be horizontal. Assume that, because of the Early effect, α increases by only one-half of 1 percent, from 0.98 to 0.985, as $|V_{CE}|$ increases from a few volts to 10 V. Then the value of β increases from $0.98/(1 - 0.98) = 49$ to $0.985/(1 - 0.985) = 66$, or about 34 percent. This numerical example illustrates that a very small change (0.5 percent) in α is reflected in a very large change (34 percent) in the value of β. It should also be clear that a slight change in α has a large effect on β, and hence upon the common-emitter curves. Therefore the common-emitter characteristics are normally subject to a wide variation even among transistors of a given type. This variability is caused by the fact that I_B is the difference between large and nearly equal currents, I_E and I_C.

EXAMPLE (*a*) Find the transistor currents in the circuit of Fig. 5-12*a*. A silicon transistor with $\beta = 100$ and $I_{CO} = 20\ nA = 2 \times 10^{-5}$ mA is under consideration. (*b*) Repeat part *a* if a 2-K emitter resistor is added to the circuit, as in Fig. 5-12*b*.

Solution *a*. Since the base is forward-biased, the transistor is not cut off. Hence it must be either in its active region or in saturation. Assume that the transistor operates in the active region. From KVL applied to the base circuit of Fig.

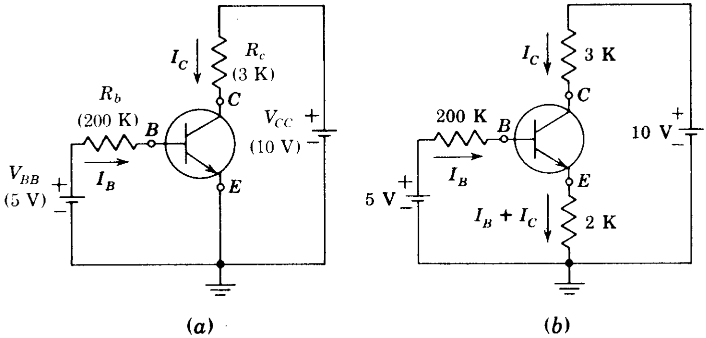


Fig. 5-12 An example illustrating how to determine whether or or not a transistor is operating in the active region.

5-12*a* (with I_B expressed in milliamperes), we have

$$-5 + 200\, I_B + V_{BE} = 0$$

As noted above, a reasonable value for V_{BE} is 0.7 V in the action region, and hence

$$I_B = \frac{5 - 0.7}{200} = 0.0215 \text{ mA}$$

Since $I_{CO} \ll I_B$, then $I_C \approx \beta I_B = 2.15$ mA.

We must now justify our assumption that the transistor is in the active region, by verifying that the collector junction is reverse-biased. From KVL applied to the collector circuit we obtain

$$-10 + 3\, I_C + V_{CB} + V_{BE} = 0$$

or

$$V_{CB} = -(3)(2.15) + 10 - 0.7 = +2.85 \text{ V}$$

For an *n-p-n* device a positive value of V_{CB} represents a reverse-biased collector junction, and hence the transistor is indeed in its active region.

Note that I_B and I_C in the active region are independent of the collector circuit resistance R_c. Hence, if R_c is increased sufficiently above 3 K, then V_{CB} changes from a positive to a negative value, indicating that the transistor is no longer in its active region. The method of calculating I_B and I_C when the transistor is in saturation is given in Sec. 5-9.

b. The current in the emitter resistor of Fig. 5-12*b* is

$$I_B + I_C \approx I_B + \beta I_B = 101\, I_B,$$

assuming $I_{CO} \ll I_B$. Applying KVL to the base circuit yields

$$-5 + 200 I_B + 0.7 + (2)(101\, I_B) = 0$$

or

$$I_B = 0.0107 \text{ mA} \qquad I_C = 100\, I_B = 1.07 \text{ mA}$$

Note that $I_{CO} = 2 \times 10^{-5}$ mA $\ll I_B$, as assumed.

To check for active circuit operation, we calculate V_{CB}. Thus

$$V_{CB} = -3I_C + 10 - (2)(101\, I_B) - 0.65$$

$$= -(3)(1.07) + 10 - (2)(101)(0.0107) - 0.7 = +3.93 \text{ V}$$

Since V_{CB} is positive, this (*n-p-n*) transistor is in its active region.

5-7 THE CE CUTOFF REGION

We might be inclined to think that cutoff in Fig. 5-10 occurs at the intersection of the load line with the current $I_B = 0$; however, we now find that appreciable collector current may exist under these conditions. From Eqs.

(5-13) and (5-14), if $I_B = 0$, then $I_E = -I_C$ and

$$I_C = -I_E = \frac{I_{CO}}{1 - \alpha} \equiv I_{CEO} \tag{5-17}$$

The actual collector current with collector junction reverse-biased and base open-circuited is designated by the symbol I_{CEO}. Since, even in the neighborhood of cutoff, α may be as large as 0.9 for germanium, then $I_C \approx 10I_{CO}$ at zero base current. Accordingly, in order to cut off the transistor, it is not enough to reduce I_B to zero. Instead, it is necessary to reverse-bias the emitter junction slightly. We shall define cutoff as the condition where the collector current is equal to the reverse saturation current I_{CO} and the emitter current is zero. It is found (Sec. 19-15) that a reverse-biasing voltage of the order of 0.1 V established across the emitter junction will ordinarily be adequate to cut off a germanium transistor. In silicon, at collector currents of the order of I_{CO}, α is very nearly zero because of recombination[5,6] in the emitter-junction transition region. Hence, even with $I_B = 0$, we find, from Eq. (5-17), that $I_C = I_{CO} = -I_E$, so that the transistor is still very close to cutoff. We verify in Sec. 19-15 that, in silicon, cutoff occurs at $V_{BE} \approx 0$ V corresponding to a base short-circuited to the emitter. *In summary, cutoff means that* $I_E = 0$, $I_C = I_{CO}$, $I_B = -I_C = -I_{CO}$, *and* V_{BE} *is a reverse voltage whose magnitude is of the order of* 0.1 V *for germanium and* 0 V *for a silicon transistor.*

The Reverse Collector Saturation Current I_{CBO} The collector current in a physical transistor (a real, nonidealized, or commercial device) when the emitter current is zero is designated by the symbol I_{CBO}. Two factors cooperate to make $|I_{CBO}|$ larger than $|I_{CO}|$. First, there exists a leakage current which flows, not through the junction, but around it and across the surfaces. The leakage current is proportional to the voltage across the junction. The second reason why $|I_{CBO}|$ exceeds $|I_{CO}|$ is that new carriers may be generated by collision in the collector-junction transition region, leading to avalanche multiplication of current and eventual breakdown. But even before breakdown is approached, this *multiplication* component of current may attain considerable proportions.

At 25°C, I_{CBO} for a germanium transistor whose power dissipation is in the range of some hundreds of milliwatts is of the order of microamperes. Under similar conditions a silicon transistor has an I_{CBO} in the range of nanoamperes. The temperature sensitivity of I_{CBO} is the same as that of the reverse saturation current I_O of a *p-n* diode (Sec. 3-5). Specifically, it is found[8] that I_{CBO} approximately doubles for every 10°C increase in temperature for both Ge and Si. However, because of the lower absolute value of I_{CBO} in silicon, these transistors may be used up to about 200°C, whereas germanium transistors are limited to about 100°C.

In addition to the variability of reverse saturation current with temperature, there is also a wide variability of reverse current among samples of a

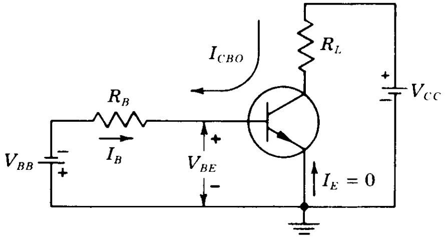


Fig. 5-13 Reverse biasing of the emitter junction to maintain the transistor in cutoff in the presence of the reverse saturation current I_{CBO} through R_B.

given transistor type. For example, the specification sheet for a Texas Instrument type 2N337 grown-diffused silicon switching transistor indicates that this type number includes units with values of I_{CBO} extending over the tremendous range from 0.2 nA to 0.3 μA. Accordingly, any particular transistor may have an I_{CBO} which differs very considerably from the average characteristic for the type.

Circuit Considerations at Cutoff Because of temperature effects, avalanche multiplication, and the wide variability encountered from sample to sample of a particular transistor type, even silicon may have values of I_{CBO} of the order of many tens of microamperes. Consider the circuit configuration of Fig. 5-13, where V_{BB} represents a biasing voltage intended to keep the transistor cutoff. Assume that the transistor is just at the point of cutoff, with $I_E = 0$, so that $I_B = -I_{CBO}$. If we require that at cutoff $V_{BE} \approx -0.1$ V, then the condition of cutoff requires that

$$V_{BE} = -V_{BB} + R_B I_{CBO} \leq -0.1 \text{ V} \tag{5-18}$$

As an extreme example consider that R_B is, say, as large as 100 K and that we want to allow for the contingency that I_{CBO} may become as large as 100 μA. Then V_{BB} must be at least 10.1 V. When I_{CBO} is small, the magnitude of the voltage across the base-emitter junction will be 10.1 V. Hence we must use a transistor whose maximum allowable reverse base-to-emitter junction voltage before breakdown exceeds 10 V. It is with this contingency in mind that a manufacturer supplies a rating for the reverse *breakdown voltage* between emitter and base, represented by the symbol BV_{EBO}. The subscript O indicates that BV_{EBO} is measured under the condition that the collector current is zero. Breakdown voltages BV_{EBO} may be as high as some tens of volts or as low as 0.5 V. If $BV_{EBO} = 1$ V, then V_{BB} must be chosen to have a maximum value of 1 V.

5-8 THE CE SATURATION REGION

In the saturation region *the collector junction (as well as the emitter junction) is forward-biased by at least the cutin voltage.* Since the voltage V_{BE} (or V_{BC})

across a forward-biased junction has a magnitude of only a few tenths of a volt, the $V_{CE} = V_{BE} - V_{BC}$ is also only a few tenths of a volt at saturation. Hence, in Fig. 5-10, the saturation region is very close to the zero-voltage axis, where all the curves merge and fall rapidly toward the origin. A load line has been superimposed on the characteristics of Fig. 5-10 corresponding to a resistance $R_L = 500\ \Omega$ and a supply voltage of 10 V. We note that in the saturation region the collector current is approximately independent of base current, for given values of V_{CC} and R_L. Hence we may consider that the onset of saturation takes place at the knee of the transistor curves in Fig. 5-10. Saturation occurs for the given load line at a base current of −0.17 mA, and at this point the collector voltage is too small to be read in Fig. 5-10. In saturation, the collector current is nominally V_{CC}/R_L, and since R_L is small, it may well be necessary to keep V_{CC} correspondingly small in order to stay within the limitations imposed by the transistor on maximum current and dissipation.

We are not able to read the collector-to-emitter saturation voltage, $V_{CE,\text{sat}}$, with any precision from the plots of Fig. 5-10. We refer instead to the characteristics shown in Fig. 5-14. In these characteristics the 0- to −0.5-V region of Fig. 5-10 has been expanded, and we have superimposed the same load line as before, corresponding to $R_L = 500\ \Omega$. We observe from Figs. 5-10 and 5-14 that V_{CE} and I_C no longer respond appreciably to base current I_B, after the base current has attained the value −0.15 mA. At this current the transistor enters saturation. For $I_B = -0.15$ mA, $|V_{CE}| \approx$ 175 mV. At $I_B = -0.35$ mA, $|V_{CE}|$ has dropped to $|V_{CE}| \approx 100$ mV. Larger magnitudes of I_B will, of course, decrease $|V_{CE}|$ slightly further.

Saturation Resistance For a transistor operating in the saturation region, a quantity of interest is the ratio $V_{CE,\text{sat}}/I_C$. This parameter is called the *common-emitter saturation resistance*, variously abbreviated R_{CS}, R_{CES}, or $R_{CE,\text{sat}}$. To specify R_{CS} properly, we must indicate the operating

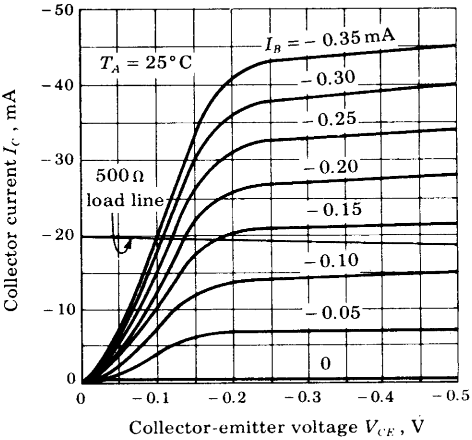


Fig. 5-14 Saturation-region common-emitter characteristics of the type 2N404 germanium transistor. A load line corresponding to V_{CC} = 10 V and $R_L = 500\ \Omega$ is superimposed. (Courtesy of Texas Instruments, Inc.)

point at which it was determined. For example, from Fig. 5-14, we find that, at $I_C = -20$ mA and $I_B = -0.35$ mA, $R_{CS} \approx -0.1/(-20 \times 10^{-3}) = 5\ \Omega$. The usefulness of R_{CS} stems from the fact, as appears in Fig. 5-14, that to the left of the knee each of the plots, for fixed I_B, may be approximated, at least roughly, by a straight line.

The Base-spreading Resistance $r_{bb'}$ Recalling that the base region is very thin (Fig. 5-5), we see that the current which enters the base region across the emitter junction must flow through a long narrow path to reach the base terminal. The cross-sectional area for current flow in the collector (or emitter) is very much larger than in the base. Hence, usually the ohmic resistance of the base is very much larger than that of the collector or emitter. The dc ohmic base resistance, designated by $r_{bb'}$, is called the *base-spreading resistance* and is of the order of magnitude of 100 Ω.

The Temperature Coefficient of the Saturation Voltages Since both junctions are forward-biased, a reasonable value for the temperature coefficient of either $V_{BE,\text{sat}}$ or $V_{BC,\text{sat}}$ is −2.5 mV/°C. In saturation the transistor consists of two forward-biased diodes back to back in series opposing. Hence it is to be anticipated that the temperature-induced voltage change in one junction will be canceled by the change in the other junction. We do indeed find such to be the case for $V_{CE,\text{sat}}$, whose temperature coefficient is about one-tenth that of $V_{BE,\text{sat}}$.

The DC Current Gain h_{FE} A transistor parameter of interest is the ratio I_C/I_B, where I_C is the collector current and I_B is the base current. This quantity is designated by β_{dc} or h_{FE}, and is known as the (negative of the) *dc beta*, the *dc forward current transfer ratio*, or the *dc current gain*.

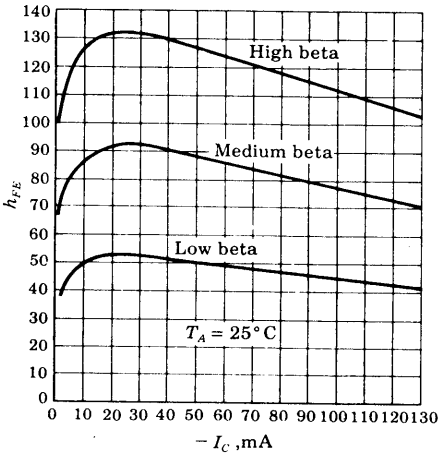


Fig. 5-15 Plots of dc current gain h_{FE} (at $V_{CE} = -0.25$ V) versus collector current for three samples of the type 2N404 germanium transistor. (Courtesy of General Electric Company.)

In the saturation region, the parameter h_{FE} is a useful number and one which is usually supplied by the manufacturer when a switching transistor is involved. We know $|I_C|$, which is given approximately by V_{CC}/R_L, and a knowledge of h_{FE} tells us the minimum base current (I_C/h_{FE}) which will be needed to saturate the transistor. For the type 2N404, the variation of h_{FE} with collector current at a low value of V_{CE} is as given in Fig. 5-15. Note the wide spread (a ratio of 3:1) in the value which may be obtained for h_{FE} even for a transistor of a particular type. Commercially available transistors have values of h_{FE} that cover the range from 10 to 150 at collector currents as small as 5 mA and as large as 30 A.

5-9 TYPICAL TRANSISTOR—JUNCTION VOLTAGE VALUES

The characteristics plotted in Fig. 5-16 of output current I_C as a function of input voltage V_{BE} for *n-p-n* germanium and silicon transistors are quite instructive and indicate the several regions of operation for a CE transistor circuit. The numerical values indicated are typical values obtained experimentally or from the theoretical equations of the following section. (The calculations are made in Sec. 19-15.) Let us examine the various portions of the transfer curves of Fig. 5-16.

The Cutoff Region *Cutoff* is defined, as in Sec. 5-5, to mean $I_E = 0$ and $I_C = I_{CO}$, and it is found that a *reverse* bias $V_{BE,\text{cutoff}} = 0.1$ V (0 V) will cut off a germanium (silicon) transistor.

What happens if a larger reverse voltage than $V_{BE,\text{cutoff}}$ is applied? It turns out that if V_E is negative and much larger than V_T, that the collector current falls slightly below I_{CO} and that the emitter current *reverses* but remains small in magnitude (less than I_{CO}).

Short-circuited Base Suppose that, instead of reverse-biasing the emitter junction, we connect the base to the emitter so that $V_E = V_{BE} = 0$. As indicated in Fig. 5-16, $I_C \equiv I_{CES}$ does not increase greatly over its cutoff value I_{CO}.

Open-circuited Base If instead of a shorted base we allow the base to "float" so that $I_B = 0$, we obtain the $I_C \equiv I_{CEO}$ given in Eq. (5-17). At low currents $\alpha \approx 0.9$ (0) for Ge (Si), and hence $I_C \approx 10 I_{CO} (I_{CO})$ for Ge (Si). The values of V_{BE} calculated for this open-base condition $(I_C = -I_E)$ are a few tens of millivolts of *forward* bias, as indicated in Fig. 5-16.

The Cutin Voltage The volt-ampere characteristic between base and emitter at constant collector-to-emitter voltage (Fig. 5-11) is not unlike the volt-ampere characteristic of a simple junction diode. When the emitter junction is reverse-biased, the base current is very small, being of the order of nanoamperes or microamperes, for silicon and germanium, respectively.

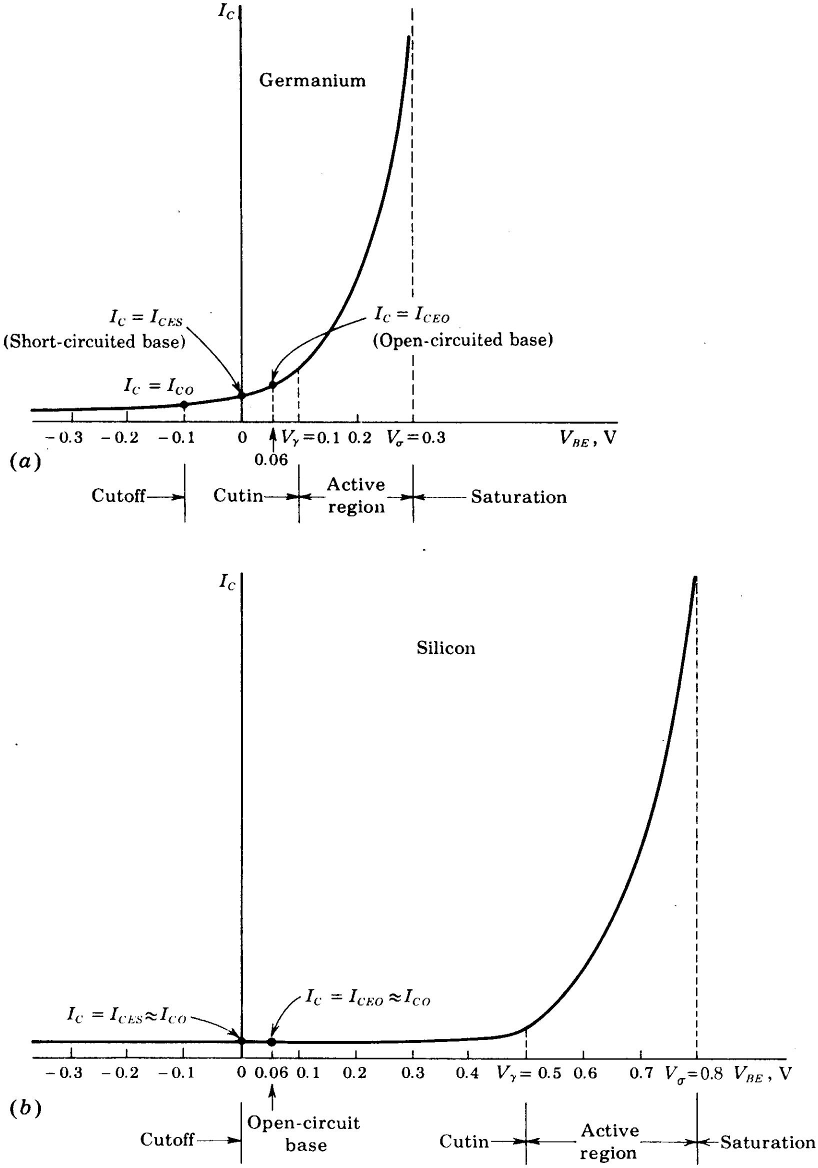


Fig. 5-16 Plots of collector current against base-to-emitter voltage for (a) germanium and (b) silicon n-p-n transistors. (I_C is not drawn to scale.)

When the emitter junction is forward-biased, again, as in the simple diode, no appreciable base current flows until the emitter junction has been forward-biased to the extent where $|V_{BE}| \geq |V_\gamma|$, where V_γ is called the *cutin voltage.* Since the collector current is nominally proportional to the base current, no appreciable collector current will flow until an appreciable base current flows. Therefore a plot of collector current against base-to-emitter voltage will exhibit a cutin voltage, just as does the simple diode.

In principle, a transistor is in its active region whenever the base-to-emitter voltage is on the forward-biasing side of the cutoff voltage, which occurs at a reverse voltage of 0.1 V for germanium and 0 V for silicon. In effect, however, a transistor enters its active region when $V_{BE} > V_\gamma$.

We may estimate the cutin voltage V_γ by assuming that $V_{BE} = V_\gamma$ when the collector current reaches, say, 1 percent of the maximum (saturation) current in the CE circuit of Fig. 5-9. Typical values of V_γ are 0.1 V for germanium and 0.5 V for silicon.

Figure 5-17 shows plots, for several temperatures, of the collector current as a function of the base-to-emitter voltage at constant collector-to-emitter voltage for a typical silicon transistor. We see that a value for V_γ of the order of 0.5 V at room temperature is entirely reasonable. The temperature dependence results from the temperature coefficient of the emitter-junction diode. Therefore the lateral shift of the plots with change in temperature and the change with temperature of the cutin voltage V_γ are approximately -2.5 mV/°C [Eq. (3-12)].

Saturation Voltages Manufacturers specify saturation values of input and output voltages in a number of different ways, in addition to supplying characteristic curves such as Figs. 5-11 and 5-14. For example, they may specify R_{CS} for several values of I_B or they may supply curves of $V_{CE,\text{sat}}$ and $V_{BE,\text{sat}}$ as functions of I_B and I_C.[9] The saturation voltages depend not only

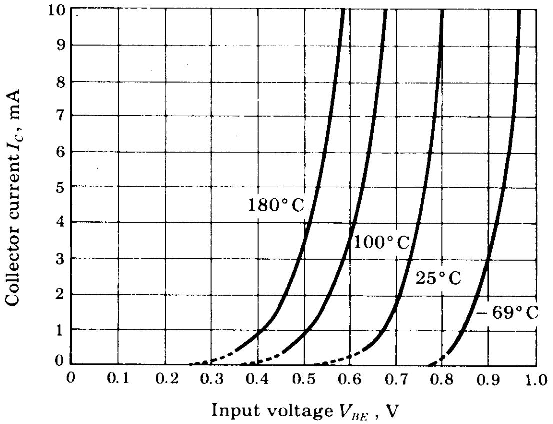


Fig. 5-17 Plot of collector current against base-to-emitter voltage for various temperatures for the type 2N337 silicon transistor. (Courtesy of Transitron Electronic Corporation.)

on the operating point, but also on the semiconductor material (germanium or silicon) and on the type of transistor construction. Typical values of saturation voltages are indicated in Table 5-1.

TABLE 5-1 Typical *n-p-n* transistor-junction voltages at 25°C

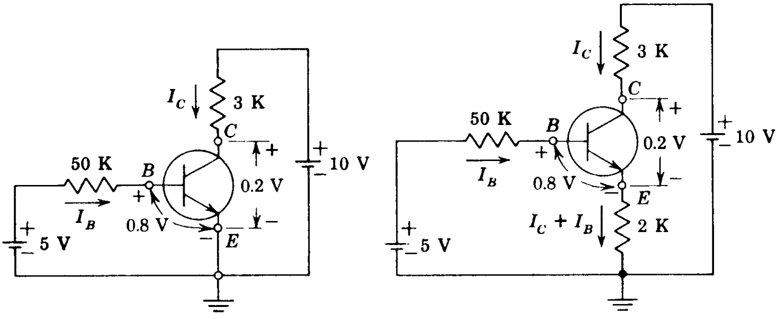
	$V_{CE,\mathrm{sat}}$	$V_{BE,\mathrm{sat}} \equiv V_\sigma$	$V_{BE,\mathrm{active}}$	$V_{BE}\dagger_{,\mathrm{cutin}} \equiv V_\gamma$	$V_{BE,\mathrm{cutoff}}$
Si	0.2	0.8	0.7	0.5	0.0
Ge	0.1	0.3	0.2	0.1	−0.1

† The temperature variation of these voltages is discussed in Secs. 5-8 and 5-9.

Summary The voltages referred to above and indicated in Fig. 5-16 are summarized in Table 5-1. The entries in the table are appropriate for an *n-p-n* transistor. For a *p-n-p* transistor the signs of all entries should be reversed. Observe that the total range of V_{BE} between cutin and saturation is rather small, being only 0.3 V. The voltage $V_{BE,\mathrm{active}}$ has been located somewhat arbitrarily, but nonetheless reasonably, near the midpoint of the active region in Fig. 5-16.

Of course, particular cases will depart from the estimates of Table 5-1. But it is unlikely that the numbers will be found in error by more than 0.1 V.

EXAMPLE (*a*) The circuits of Fig. 5-12*a* and *b* are modified by changing the base-circuit resistance from 200 to 50 K (as indicated in Fig. 5-18). If $h_{FE} = 100$, determine whether or not the silicon transistor is in saturation and find I_B and I_C. (*b*) Repeat with the 2K emitter resistance added.



(*a*) (*b*)

Fig. 5-18 An example illustrating how to determine whether or not a transistor is operating in the saturation region.

Solution Assume that the transistor is in saturation. Using the values $V_{BE,\text{sat}}$ and $V_{CE,\text{sat}}$ in Table 5-1, the circuit of Fig. 5-18*a* is obtained. Applying KVL to the base circuit gives

$$-5 + 50\,I_B + 0.8 = 0$$

or

$$I_B = \frac{4.2}{50} = 0.084 \text{ mA}$$

Applying KVL to the collector circuit yields

$$-10 + 3\,I_C + 0.2 = 0$$

or

$$I_C = \frac{9.8}{3} = 3.27 \text{ mA}$$

The minimum value of base current required for saturation is

$$(I_B)_{\text{min}} = \frac{I_C}{h_{FE}} = \frac{3.27}{100} = 0.033 \text{ mA}$$

Since $I_B = 0.084 > I_{B,\text{min}} = 0.033$ mA, we have verified that the transistor is in saturation.

b. If the 2-K emitter resistance is added, the circuit becomes that in Fig. 5-18*b*. Assume that the transistor is in saturation. Applying KVL to the base and collector circuits, we obtain

$$-5 + 50\,I_B + 0.8 + 2\,(I_C + I_B) = 0$$

$$-10 + 3\,I_C + 0.2 + 2\,(I_C + I_B) = 0$$

If these simultaneous equations are solved for I_C and I_B, we obtain

$$I_C = 1.95 \text{ mA} \qquad I_B = 0.0055 \text{ mA}$$

Since $(I_B)_{\text{min}} = I_C/h_{FE} = 0.0195 \text{ mA} > I_B = 0.0055$, the transistor is *not* in saturation. Hence the device must be operating in the active region. Proceeding exactly as we did for the circuit of Fig. 5-12*b* (but with the 200 K replaced by 50 K), we obtain

$$I_C = 1.71 \text{ mA} \qquad I_B = 0.0171 \text{ mA} = 17\ \mu\text{A} \qquad V_{CB} = 0.72 \text{ V}$$

5-10 COMMON-EMITTER CURRENT GAIN

Three different definitions of current gain appear in the literature. The interrelationships between these are now to be found.

Large-signal Current Gain β We define β in terms of α by Eq. (5-15). From Eq. (5-16), with I_{CO} replaced by I_{CBO}, we find

$$\beta = \frac{I_C - I_{CBO}}{I_B - (-I_{CBO})} \tag{5-19}$$

In Sec. 5-7 we define *cutoff* to mean that $I_E = 0$, $I_C = I_{CBO}$, and $I_B = -I_{CBO}$. Consequently, Eq. (5-19) gives the ratio of the collector-current increment to the base-current change from cutoff to I_B, and hence β *represents the* (negative of the) *large-signal current gain of a common-emitter transistor.* This parameter is of primary importance in connection with the biasing and stability of transistor circuits, as discussed in Chap. 9.

DC Current Gain h_{FE} In Sec. 5-8 we define the dc current gain by

$$\beta_{\mathrm{dc}} \equiv \frac{I_C}{I_B} \equiv h_{FE} \tag{5-20}$$

In that section it is noted that h_{FE} is most useful in connection with determining whether or not a transistor is in saturation. In general, the base current (and hence the collector current) is large compared with I_{CBO}. Under these conditions the large-signal and the dc betas are approximately equal; then $h_{FE} \approx \beta$.

Small-signal Current Gain h_{fe} We define β' as the ratio of a collector-current increment ΔI_C for a small base-current change ΔI_B (at a given quiescent operating point, at a fixed collector-to-emitter voltage V_{CE}), or

$$\beta' \equiv \left.\frac{\partial I_C}{\partial I_B}\right|_{V_{CE}} = h_{fe} \tag{5-21}$$

Clearly, β' is (the negative of) the *small-signal* current gain. If β were independent of current, we see from Eq. (5-20) that $\beta' = \beta \approx h_{FE}$. However, Fig. 5-15 indicates that β is a function of current, and differentiating Eq. (5-16) with respect to I_C gives (with $I_{CO} = I_{CBO}$)

$$1 = (I_{CBO} + I_B)\frac{\partial \beta}{\partial I_C} + \beta\frac{\partial I_B}{\partial I_C} \tag{5-22}$$

The small-signal CE gain β' is used in the analysis of small-signal amplifier circuits and is designated by h_{fe} in Chap. 8. Using Eq. (5-21), and with $\beta' = h_{fe}$ and $\beta = h_{FE}$, Eq. (5-22) becomes

$$h_{fe} = \frac{h_{FE}}{1 - (I_{CBO} + I_B)(\partial h_{FE}/\partial I_C)} \tag{5-23}$$

Since h_{FE} versus I_C given in Fig. 5-15 shows a maximum, h_{fe} is larger than h_{FE} for small currents (to the left of the maximum) and h_{fe} is smaller than h_{FE} for currents larger than that corresponding to the maximum. Over most of the wide current range in Fig. 5-14, h_{fe} differs from h_{FE} by less than 20 percent.

It should be emphasized that Eq. (5-23) is valid in the active region only. From Fig. 5-14 we see that $h_{fe} \to 0$ in the saturation region because $\Delta I_C \to 0$ for a small increment ΔI_B.

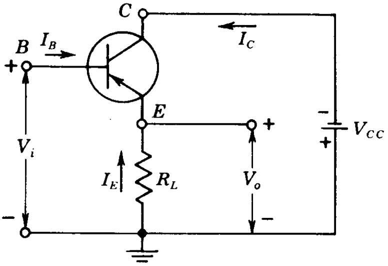


Fig. 5-19 The transistor common-collector configuration.

5-11 THE COMMON-COLLECTOR CONFIGURATION

Another transistor-circuit configuration, shown in Fig. 5-19, is known as the common-collector configuration. The circuit is basically the same as the circuit of Fig. 5-9, with the exception that the load resistor is in the emitter lead rather than in the collector circuit. If we continue to specify the operation of the circuit in terms of the currents which flow, the operation for the common-collector is much the same as for the common-emitter configuration. When the base current is I_{CO}, the emitter current will be zero, and no current will flow in the load. As the transistor is brought out of this back-biased condition by increasing the magnitude of the base current, the transistor will pass through the active region and eventually reach saturation. In this condition all the supply voltage, except for a very small drop across the transistor, will appear across the load.

5-12 ANALYTICAL EXPRESSIONS FOR TRANSISTOR CHARACTERISTICS

The dependence of the currents in a transistor upon the junction voltages, or vice versa, may be obtained by starting with Eq. (5-6), repeated here for convenience:

$$I_C = -\alpha_N I_E - I_{CO}(\epsilon^{V_C/V_T} - 1) \tag{5-24}$$

We have added the subscript N to α to indicate that we are using the transistor in the *normal* manner. We must recognize, however, that there is no essential reason which constrains us from using a transistor in an *inverted* fashion, that is, interchanging the roles of the emitter junction and the collector junction. From a practical point of view, such an arrangement might not be as effective as the *normal* mode of operation, but this matter does not concern us now. With this inverted mode of operation in mind, we may now write, in correspondence with Eq. (5-24),

$$I_E = -\alpha_I I_C - I_{EO}(\epsilon^{V_E/V_T} - 1) \tag{5-25}$$

Here α_I is the *inverted* common-base current gain, just as α_N in Eq. (5-24) is the current gain in *normal* operation. I_{EO} is the emitter-junction reverse saturation current, and V_E is the voltage drop from p side to n side at the emitter junction and is positive for a forward-biased emitter.

Equations (5-24) and (5-25) were derived in a heuristic manner. A physical analysis of the transistor currents by Ebers and Moll[7] verifies these equations (Sec. 19-13). This quantitative study reveals that the parameters α_N, α_I, I_{CO}, and I_{EO} are not independent, but are related by the condition

$$\alpha_I I_{CO} = \alpha_N I_{EO} \tag{5-26}$$

Manufacturer's data sheets often provide information about α_N, I_{CO}, and I_{EO}, so that α_I may be determined. For many transistors I_{EO} lies in the range $0.5I_{CO}$ to I_{CO}.

Since the sum of the three currents must be zero, the base current is given by

$$I_B = -(I_E + I_C) \tag{5-27}$$

If three of the four parameters α_N, α_I, I_{CO}, and I_{EO} are known, the equations in this section allow calculations of the three currents for given values of junction voltages V_C and V_E. Explicit expressions for I_C and I_E in terms of V_C and V_E are found in Sec. 19-13.

In the literature, α_R (*reversed* alpha) and α_F (*forward* alpha) are sometimes used in place of α_I and α_N, respectively.

Reference Polarities The symbol $V_C(V_E)$ represents the drop across the collector (emitter) junction and is positive if the junction is forward-biased. The reference directions for currents and voltages are indicated in Fig. 5-20. Since V_{CB} represents the voltage drop from collector-to-base terminals, then V_{CB} differs from V_C by the ohmic drop in the base-spreading resistance $r_{bb'}$, or

$$V_{CB} = V_C - I_B r_{bb'} \tag{5-28}$$

The Ebers-Moll Model Equations (5-24) and (5-25) have a simple interpretation in terms of a circuit known as the *Ebers-Moll model.*[7] This model is shown in Fig. 5-21 for a *p-n-p* transistor. We see that it involves two ideal diodes placed back to back with reverse saturation currents $-I_{EO}$ and $-I_{CO}$ and two dependent current-controlled current sources shunting the ideal diodes. For a *p-n-p* transistor, both I_{CO} and I_{EO} are negative, so that $-I_{CO}$ and $-I_{EO}$ are positive values, giving the magnitudes of the reverse saturation currents of the diodes. The current sources account for the minority-carrier transport across the base. An application of KCL to the collector node of Fig. 5-21 gives

$$I_C = -\alpha_N I_E + I = -\alpha_N I_E + I_o(\epsilon^{V_C/V_T} - 1)$$

where the diode current I is given by Eq. (3-9). Since I_o is the magnitude

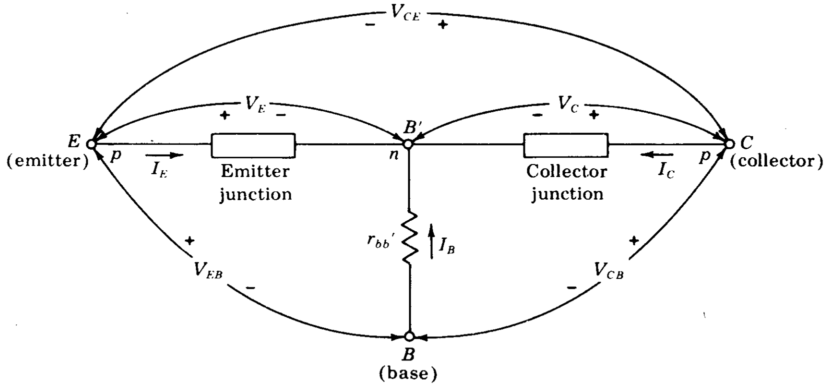


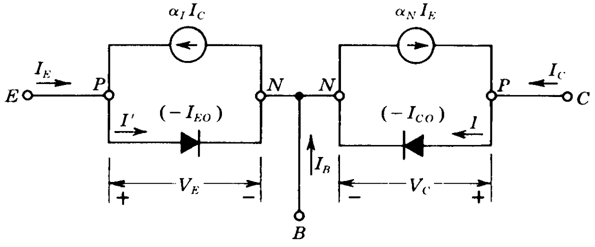
Fig. 5-20 Defining the voltages and currents used in the Ebers-Moll equations. For either a p-n-p or an n-p-n transistor, a positive value of current means that positive charge flows into the junction and a positive $V_E(V_C)$ means that the emitter (collector) junction is forward-biased (the p side positive with respect to the n side).

of the reverse saturation, then $I_o = -I_{CO}$. Substituting this value of I_o into the preceding equation for I_C yields Eq. (5-24).

This model is valid for both forward and reverse static voltages applied across the transistor junctions. It should be noted that we have omitted the base-spreading resistance from Fig. 5-20 and have neglected the difference between I_{CBO} and I_{CO}.

Observe from Fig. 5-21 that the dependent current sources can be eliminated from this figure provided $\alpha_N = \alpha_I = 0$. For example, by making the base width much larger than the diffusion length of minority carriers in the base, all minority carriers will recombine in the base and none will survive to reach the collector. For this case the current gain α will be zero. Under these conditions, transistor action ceases, and we simply have two diodes placed back to back. This discussion shows why it is impossible to construct a transistor by simply connecting two separate (isolated) diodes in series opposing. A cascade of two p-n diodes exhibits transistor properties (for

Fig. 5-21 The Ebers-Moll model for a p-n-p transistor.



example, it is capable of amplification) only if carriers injected across one junction *diffuse* across the second junction.

Voltages as Functions of Currents We may solve explicitly for the junction voltages in terms of the currents from Eqs. (5-24) and (5-25), with the result that

$$V_E = V_T \ln\left(1 - \frac{I_E + \alpha_I I_C}{I_{EO}}\right) \tag{5-29}$$

$$V_C = V_T \ln\left(1 - \frac{I_C + \alpha_N I_E}{I_{CO}}\right) \tag{5-30}$$

We now derive the analytic expression for the common-emitter characteristics of Fig. 5-10. The abscissa in this figure is the collector-to-emitter voltage $V_{CE} = V_E - V_C$ for an *n-p-n* transistor and is $V_{CE} = V_C - V_E$ for a *p-n-p* transistor (remember that V_C and V_E are positive at the *p* side of the junction). Hence the common-emitter characteristics are found by subtracting Eqs. (5-29) and (5-30) and by eliminating I_E by the use of Eq. (5-27). The resulting equation can be simplified provided that the following inequalities are valid: $I_B \gg I_{EO}$ and $I_B \gg I_{CO}/\alpha_N$. After some manipulations and by the use of Eqs. (5-15) and (5-26), we obtain (except for very small values of I_B)

$$V_{CE} = \pm V_T \ln \frac{\dfrac{1}{\alpha_I} + \dfrac{1}{\beta_I}\dfrac{I_C}{I_B}}{1 - \dfrac{1}{\beta}\dfrac{I_C}{I_B}} \tag{5-31}$$

where

$$\beta_I \equiv \frac{\alpha_I}{1 - \alpha_I} \qquad \text{and} \qquad \beta_N \equiv \beta \equiv \frac{\alpha}{1 - \alpha} \tag{5-32}$$

Note that the + sign in Eq. (5-31) is used for an *n-p-n* transistor, and the − sign for a *p-n-p* device. For a *p-n-p* germanium-type transistor, at $I_C = 0$, $V_{CE} = -V_T \ln (1/\alpha_I)$, so that the *common-emitter characteristics do not pass through the origin.* For $\alpha_I = 0.78$ and $V_T = 0.026$ V, we have $V_{CE} \approx -6$ mV at room temperature. This voltage is so small that the curves of Fig. 5-10 look as if they pass through the origin, but they are actually displaced to the right by a few millivolts.

If I_C is increased, then V_{CE} rises only slightly until I_C/I_B approaches β. For example, even for $I_C/I_B = 0.9\beta = 90$ (for $\beta = 100$),

$$V_{CE} = -0.026 \ln \frac{1/0.78 + 90/3.5}{1 - 0.9} \approx -0.15 \text{ V}$$

This voltage can barely be detected at the scale to which Fig. 5-10 is drawn, and hence near the origin it appears as if the curves rise vertically. However, note that Fig. 5-14 confirms that a voltage of the order of 0.2 V is required for I_C to reach 0.9 of its maximum value.

The maximum value of I_C/I_B is β, and as this value of I_C/I_B is approached,

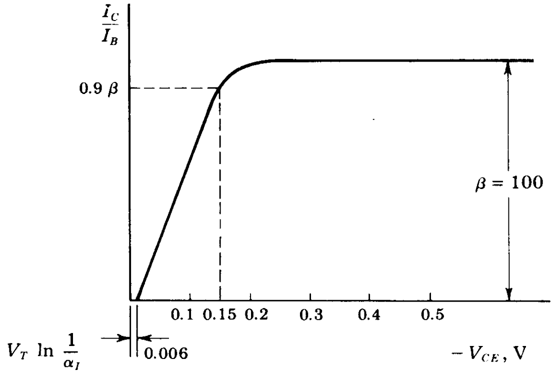


Fig. 5-22 The common-emitter output characteristic for a *p-n-p* transistor as obtained analytically.

$V_{CE} \to -\infty$. Hence, as I_C/I_B increases from 0.9β to β, $|V_{CE}|$ increases from 0.15 V to infinity. A plot of the theoretical common-emitter characteristic is indicated in Fig. 5-22. We see that, at a fixed value of V_{CE}, the ratio I_C/I_B is a constant. Hence, for equal increments in I_B, we should obtain equal increments in I_C at a given V_{CE}. This conclusion is fairly well satisfied by the curves in Fig. 5-10. However, the $I_B = 0$ curve seems to be inconsistent since, for a constant I_C/I_B, this curve should coincide with the $I_C = 0$ axis. This discrepancy is due to the approximation made in deriving Eq. (5-31), which is not valid for $I_B = 0$.

The theoretical curve of Fig. 5-22 is much flatter than the curves of Fig. 5-10 because we have implicitly assumed that α_N is truly constant. As already pointed out, a very slight increase of α_N with V_{CE} can account for the slopes of the common-emitter characteristic.

5-13 MAXIMUM VOLTAGE RATING[9]

Even if the rated dissipation of a transistor is not exceeded, there is an upper limit to the maximum allowable collector-junction voltage since, at high voltages, there is the possibility of voltage breakdown in the transistor. Two types of breakdown are possible, *avalanche breakdown*, discussed in Sec. 3-11, and *reach-through*, discussed below.

Avalanche Multiplication The maximum reverse-biasing voltage which may be applied before breakdown between the collector and base terminals of the transistor, under the condition that the emitter lead be open-circuited, is represented by the symbol BV_{CBO}. This breakdown voltage is a characteristic of the transistor alone. Breakdown may occur because of avalanche multiplication of the current I_{CO} that crosses the collector junction. As a result of this multiplication, the current becomes MI_{CO}, in which M is the factor by which the original current I_{CO} is multiplied by the avalance effect. (We neglect leakage current, which does not flow through the junction and is therefore not subject to avalanche multiplication.) At a high enough voltage, namely, BV_{CBO}, the multiplication factor M becomes nominally infinite, and the region of breakdown is then attained. Here the current rises abruptly, and large changes in current accompany small changes in applied voltage.

The avalanche multiplication factor depends on the voltage V_{CB} between collector and base. We shall consider that

$$M \equiv \frac{1}{1 - (V_{CB}/BV_{CBG})^n} \tag{5-33}$$

Equation (5-33) is employed because it is a simple expression which gives a good empirical fit to the breakdown characteristics of many transistor types. The parameter n is found to be in the range of about 2 to 10, and controls the sharpness of the onset of breakdown.

If a current I_E is caused to flow across the emitter junction, then, neglecting the avalanche effect, a fraction αI_E, where α is the common-base current gain, reaches the collector junction. Taking multiplication into account, I_C has the magnitude $M\alpha I_E$. Consequently, it appears that, in the presence of avalanche multiplication, the transistor behaves as though its common-base current gain were $M\alpha$.

An analysis[9] of avalanche breakdown for the CE configuration indicates that the collector-to-emitter breakdown voltage *with open-circuited base*, designated BV_{CEO}, is

$$BV_{CEO} = BV_{CBO}\sqrt[n]{\frac{1}{h_{FE}}} \tag{5-34}$$

For an *n-p-n* germanium transistor, a reasonable value for n, determined experimentally, is $n = 6$. If we now take $h_{FE} = 50$, we find that

$$BV_{CEO} = 0.52BV_{CBO}$$

so that if $BV_{CBO} = 40$ V, BV_{CEO} is about half as much, or about 20 V. Ideal-

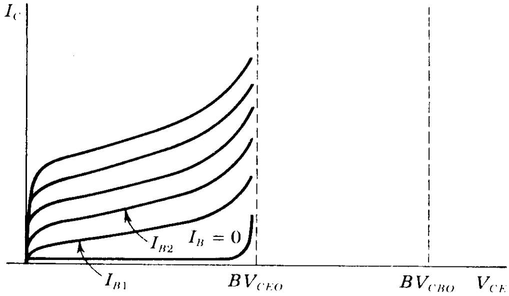


Fig. 5-23 Idealized common-emitter characteristics extended into the breakdown region.

ized common-emitter characteristics extended into the breakdown region are shown in Fig. 5-23. If the base is not open-circuited, these breakdown characteristics are modified, the shapes of the curves being determined by the base-circuit connections. In other words, the maximum allowable collector-to-emitter voltage depends not only upon the transistor, but also upon the circuit in which it is used.

Reach-through The second mechanism by which a transistor's usefulness may be terminated as the collector voltage is increased is called *punch-through*, or *reach-through*, and results from the increased width of the collector-junction transition region with increased collector-junction voltage (the Early effect).

The transition region at a junction is the region of uncovered charges on both sides of the junction at the positions occupied by the impurity atoms. As the voltage applied across the junction increases, the transition region penetrates deeper into the base (Fig. 5-8*a*). Since the base is very thin, it is possible that, at moderate voltages, the transition region will have spread completely across the base to reach the emitter junction, as indicated in Fig. 5-24, which should be compared with Fig. 5-8. The emitter barrier is now V', which is smaller than the normal value $V_o - |V_{EB}|$ because the collector voltage has "reached through" the base region. This lowering of the emitter-junction voltage may result in an excessively large emitter current, thus placing an upper limit on the magnitude of the collector voltage.

Punch-through differs from avalanche breakdown in that it takes place at a fixed voltage [given by V_j in Eq. (3-21), with $W = W_B$] between collector and base, and is not dependent on circuit configuration. In a particular transistor, the voltage limit is determined by punch-through or breakdown, whichever occurs at the lower voltage.

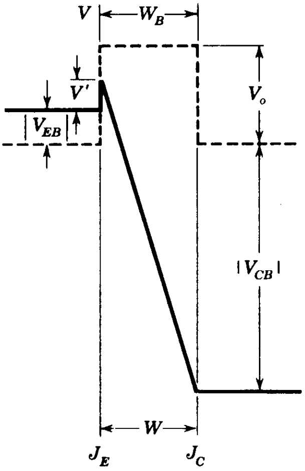


Fig. 5-24 The potential variation through a *p-n-p* transistor after "reach-through" when the effective base width (Fig. 5-8) $W'_B = W_B - W$ has been reduced to zero. The effective emitter barrier is V'.

5-14 THE PHOTOTRANSISTOR

The phototransistor (also called *photoduodiode*) is a much more sensitive semiconductor photodevice than the *p-n* photodiode. The phototransistor is usually connected in a common-emitter configuration with the base open, and radiation is concentrated on the region near the collector junction J_C, as in Fig. 5-25*a*. The operation of this device can be understood if we recognize that the junction J_E is slightly forward-biased (Fig. 5-16, open-circuited base), and the junction J_C is reverse-biased (that is, the transistor is biased in the active region). Assume, first, that there is no radiant excitation. Under these circumstances minority carriers are generated thermally, and the electrons crossing from the base to the collector, as well as the holes crossing from the collector to the base, constitute the reverse saturation collector current I_{CO} (Sec. 5-2). The collector current is given by Eq. (5-16), with $I_B = 0$; namely,

$$I_C = (\beta + 1)I_{CO} \tag{5-35}$$

If the light is now turned on, additional minority carriers are photogenerated, and these contribute to the reverse saturation current in exactly the same manner as do the thermally generated minority charges. If the component of the reverse saturation current due to the light is designated I_L, the total collector current is

$$I_C = (\beta + 1)(I_{CO} + I_L) \tag{5-36}$$

We note that, due to transistor action, the current caused by the radiation is multiplied by the large factor $\beta + 1$.

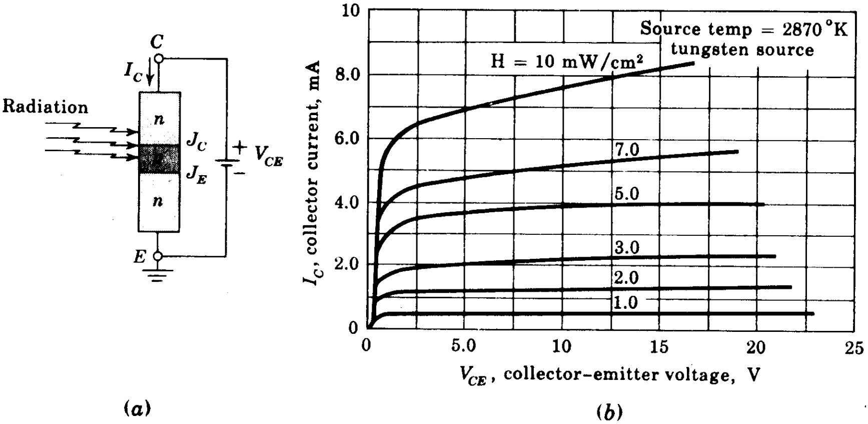


Fig. 5-25 (a) A phototransistor. (b) The output characteristics of the MRD 450 *n-p-n* silicon phototransistor. (Courtesy of Motorola Semiconductor Products, Inc.)

Typical volt-ampere characteristics are shown in Fig. 5-23*b* for an *n-p-n* planar phototransistor for different values of illumination intensities. Note the similarity between this family of curves and those in Fig. 5-10 for the CE transistor output characteristics with base current (instead of illumination) as a parameter. It is also possible to bring out the base lead and to inject a base current I_B. The current I_C in Eq. (5-36) is then increased by the term βI_B.

REFERENCES

1. Shockley, W.: The Theory of *p-n* Junctions in Semiconductors and *p-n* Junction Transistors, *Bell System Tech. J.*, vol. 28, pp. 435–489, July, 1949.

   Middlebrook, R. D.: "An Introduction to Junction Transistor Theory," pp. 115–130, John Wiley & Sons, Inc., New York, 1957.

   Terman, F. E.: "Electronic and Radio Engineering," 4th ed., pp. 747–760, McGraw-Hill Book Company, New York, 1955.

   Moll, J. L.: Junction Transistor Electronics, *Proc. IRE*, vol. 43, pp. 1807–1819, December, 1955.

2. Phillips, A. B.: "Transistor Engineering," chap. 1, McGraw-Hill Book Company, New York, 1962.

3. Texas Instruments, Inc.: J. Miller (ed.), "Transistor Circuit Design," chap. 1, McGraw-Hill Book Company, New York, 1963.

4. Early, J. M.: Effects of Space-charge Layer Widening in Junction Transistors, *Proc. IRE*, vol. 40, pp. 1401–1406, November, 1952.

5. Sah, C. T., R. N. Noyce, and W. Shockley: Carrier-generation and Recombination in *p-n* Junctions and *p-n* Junction Characteristics, *Proc. IRE*, vol. 45, pp. 1228–1243, September, 1957.

   Pritchard, R. L.: Advances in the Understanding of the P-N Junction Triode, *Proc. IRE*, vol. 46, pp. 1130–1141, June, 1958.

6. Ref. 2, pp. 236–237.

7. Ebers, J. J., and J. L. Moll: Large-signal Behavior of Junction Transistors, *Proc. IRE*, vol. 42, pp. 1761–1772, December, 1954.

8. Millman, J., and H. Taub: "Pulse, Digital, and Switching Waveforms," p. 196, McGraw-Hill Book Company, New York, 1965.

9. Ref. 8, chap. 6.

REVIEW QUESTIONS

5-1 Draw the circuit symbol for a *p-n-p* transistor and indicate the reference directions for the three currents and the reference polarities for the three voltages.

5-2 Repeat Rev. 5-1 for an *n-p-n* transistor.

5-3 For a *p-n-p* transistor biased in the active region, plot (in each region E, B, and C) (*a*) the potential variation; (*b*) the minority-carrier concentration. (*c*) Explain the shapes of the plots in (*a*) and (*b*).

5-4 (*a*) For a *p-n-p* transistor biased in the active region, indicate the various electron and hole current components crossing each junction and entering (or leaving) the base terminal. (*b*) Which of the currents is proportional to the gradient of p_n at J_E and J_C, respectively? (*c*) Repeat part *b* with p_n replaced by n_p. (*d*) What is the physical origin of the several current components crossing the base terminal?

5-5 (*a*) From the currents indicated in Rev. 5-4 obtain an expression for the collector current I_C. Define each symbol in this equation. (*b*) Generalize the equation for I_C in part *a* so that it is valid even if the transistor is not operating in its active region.

5-6 (*a*) Define the *current gain* α in words and as an equation. (*b*) Repeat part *a* for the parameter α'.

5-7 Describe the fabrication of an alloy transistor.

5-8 For a *p-n-p* transistor in the active region, what is the sign (positive or negative) of I_E, I_C, I_B, V_{CB}, and V_{EB}?

5-9 Repeat Rev. 5-8 for an *n-p-n* transistor.

5-10 (*a*) Sketch a family of CB output characteristics for a transistor. (*b*) Indicate the active, cutoff, and saturation regions. (*c*) Explain the shapes of the curves qualitatively.

5-11 (*a*) Sketch a family of CB input characteristics for a transistor. (*b*) Explain the shapes of the curves qualitatively.

5-12 Explain *base-width modulation* (the Early effect) with the aid of plots of potential and minority concentration throughout the base region.

5-13 Explain qualitatively the three consequences of base-width modulation.

5-14 Define the following regions in a transistor: (*a*) active; (*b*) saturation; (*c*) cutoff.

5-15 (*a*) Draw the circuit of transistor in the CE configuration. (*b*) Sketch the output characteristics. (*c*) Indicate the active, saturation, and cutoff regions.

5-16 (*a*) Sketch a family of CE input characteristics. (*b*) Explain the shape of these curves qualitatively.

5-17 (*a*) Derive the expression for I_C versus I_B for a CE transistor configuration in the active region. (*b*) For $I_B = 0$, what is I_C?

5-18 (*a*) What is the order of magnitude of the reverse collector saturation current I_{CBO} for a silicon transistor? (*b*) How does I_{CBO} vary with temperature?

5-19 Repeat Rev. 5-18 for a germanium transistor.

5-20 Why does I_{CBO} differ from I_{CO}?

5-21 (*a*) Define *saturation resistance* for a CE transistor. (*b*) Give its order of magnitude.

5-22 (*a*) Define *base-spreading resistance* for a transistor. (*b*) Give its order of magnitude.

5-23 What is the order of magnitude of the temperature coefficients of $V_{BE,\text{sat}}$, $V_{BC,\text{sat}}$, and $V_{CE,\text{sat}}$?

5-24 (*a*) Define h_{FE}. (*b*) Plot h_{FE} versus I_C.

5-25 (*a*) Give the order of magnitude of V_{BE} at cutoff for a silicon transistor. (*b*) Repeat part *a* for a germanium transistor. (*c*) Repeat parts *a* and *b* for the cutin voltage.

5-26 Is $|V_{BE,\mathrm{sat}}|$ greater or less than $|V_{CE,\mathrm{sat}}|$? Explain.

5-27 (*a*) What is the range in volts for V_{BE} between cutin and saturation for a silicon transistor? (*b*) Repeat part *a* for a germanium transistor.

5-28 What is the collector current relative to I_{CO} in a silicon transistor (*a*) if the base is shorted to the emitter? (*b*) If the base floats? (*c*) Repeat parts *a* and *b* for a germanium transistor.

5-29 Consider a transistor circuit with resistors R_b, R_c, and R_e in the base, collector, and emitter legs, respectively. The biasing voltages are V_{BB} and V_{CC} in base and collector circuits, respectively. (*a*) Outline the method for finding the quiescent currents, assuming that the transistor operates in the active region. (*b*) How do you test to see if your assumption is correct?

5-30 Repeat Rev. 5-29, assuming that the transistor is in saturation.

5-31 For a CE transistor define (in words and symbols) (*a*) β; (*b*) $\beta_{dc} = h_{FE}$; (*c*) $\beta' = h_{fe}$.

5-32 Derive the relationship between h_{FE} and h_{fe}.

5-33 (*a*) For what condition is $\beta \approx h_{FE}$? (*b*) For what condition is $h_{FE} \approx h_{fe}$?

5-34 (*a*) Draw the circuit of a CC transistor configuration. (*b*) Indicate the input and output terminals.

5-35 What is meant by the *inverted mode* of operation of a transistor?

5-36 (*a*) Write the Ebers and Moll equations. (*b*) Sketch the circuit model which satisfies these equations.

5-37 Discuss the two possible sources of breakdown in a transistor as the collector-to-emitter voltage is increased.

5-38 (*a*) Sketch the circuit of a phototransistor. (*b*) The radiation is concentrated near which junction? Explain why. (*c*) Describe the physical action of this device.

5-39 (*a*) For a phototransistor in the active region, write the expression for the collector current. Define all terms. (*b*) Sketch the family of output characteristics.

7 / INTEGRATED CIRCUITS: FABRICATION AND CHARACTERISTICS

An integrated circuit consists of a single-crystal chip of silicon, typically 50 by 50 mils in cross section, containing both active and passive elements and their interconnections. Such circuits are produced by the same processes used to fabricate individual transistors and diodes. These processes include epitaxial growth, masked impurity diffusion, oxide growth, and oxide etching, using photolithography for pattern definition. A method of batch processing is employed which offers excellent repeatability and is adaptable to the production of large numbers of integrated circuits at low cost. In this chapter we describe the basic processes involved in fabricating an integrated circuit.

7-1 INTEGRATED–CIRCUIT TECHNOLOGY

The fabrication of integrated circuits is based on materials, processes, and design principles which constitute a highly developed semiconductor (planar-diffusion) technology. The basic structure of an integrated circuit is shown in Fig. 7-1*b*, and consists of four distinct layers of material. The bottom layer ① (6 mils thick) is *p*-type silicon and serves as a *substrate* upon which the integrated circuit is to be built. The second layer ② is thin (typically 25 μm = 1 mil) *n*-type material which is grown as a single-crystal extension of the substrate. All active and passive components are built within the thin *n*-type layer using a series of diffusion steps. These components are transistors, diodes, capacitors, and resistors, and they are made by diffusing *p*-type and *n*-type impurities. The most complicated component fabricated is the transistor, and all other elements are constructed with one or more of the processes required to make a transistor. In the fabrication of all the above elements it is necessary to distribute impurities in certain precisely defined regions within the second (*n*-type) layer. The selective diffusion of impurities is accomplished by using SiO_2 as a barrier which protects portions of the wafer against impurity penetration. Thus the third layer of material ③ is silicon dioxide, and it also provides protection of the semiconductor surface against contamination. In the regions where diffusion is to take place, the SiO_2 layer is etched away, leaving the rest of the wafer protected against diffusion. To permit selective etching, the SiO_2 layer must be subjected to a photolithographic process, described in Sec. 7-4. Finally, a fourth metallic (aluminum) layer ④ is added to supply the necessary interconnections between components.

The *p*-type substrate which is required as a foundation for the integrated circuit is obtained by growing an ingot (1 to 2 in. in diameter and about 10 in. long) from a silicon melt with a predetermined number of impurities. The crystal ingot is subsequently sliced into round wafers approximately 6 mils thick, and one side of each wafer is lapped and polished to eliminate surface imperfections.

We are now in a position to appreciate some of the significant advantages of the integrated-circuit technology. Let us consider a 1-in.-square wafer divided into 400 chips of surface area 50 mil by 50 mil. We demonstrate in this chapter that a reasonable area under which a component (say, a transistor) is fabricated is 50 mils^2. Hence each chip (each integrated circuit) contains 50 separate components, and there are $50 \times 400 = 20{,}000$ components/in.^2 on each wafer.

If we process 10 wafers in a batch, we can manufacture 4,000 integrated circuits simultaneously, and these contain 200,000 components. Some of the chips will contain faults due to imperfections in the manufacturing process, but the *yield* (the percentage of fault-free chips per wafer) is extremely large.

The following advantages are offered by integrated-circuit technology as compared with discrete components interconnected by conventional techniques:

1. Low cost (due to the large quantities processed).
2. Small size.
3. High reliability. (All components are fabricated simultaneously, and there are no soldered joints.)
4. Improved performance. (Because of the low cost, more complex circuitry may be used to obtain better functional characteristics.)

In the next sections we examine the processes required to fabricate an integrated circuit.

7-2 BASIC MONOLITHIC INTEGRATED CIRCUITS[1,2]

We now examine in some detail the various techniques and processes required to obtain the circuit of Fig. 7-1*a* in an integrated form, as shown in Fig. 7-1*b*. This configuration is called a monolithic integrated circuit because it is formed on a single silicon chip. The word "monolithic" is derived from the Greek *monos*, meaning "single," and *lithos*, meaning "stone." Thus a monolithic circuit is built into a single stone, or single crystal.

In this section we describe qualitatively a complete epitaxial-diffused fabrication process for integrated circuits. In subsequent sections we examine in more detail the epitaxial, photographic, and diffusion processes involved. The logic circuit of Fig. 7-1*a* is chosen for discussion because it contains typical components: a resistor, diodes, and a transistor. These elements (and also capacitors with small values of capacitances) are the components encountered in integrated circuits. The monolithic circuit is formed by the steps indicated in Fig. 7-2 and described below.

Step 1. Epitaxial Growth An *n*-type epitaxial layer, typically 25 μm thick, is grown onto a *p*-type substrate which has a resistivity of typically

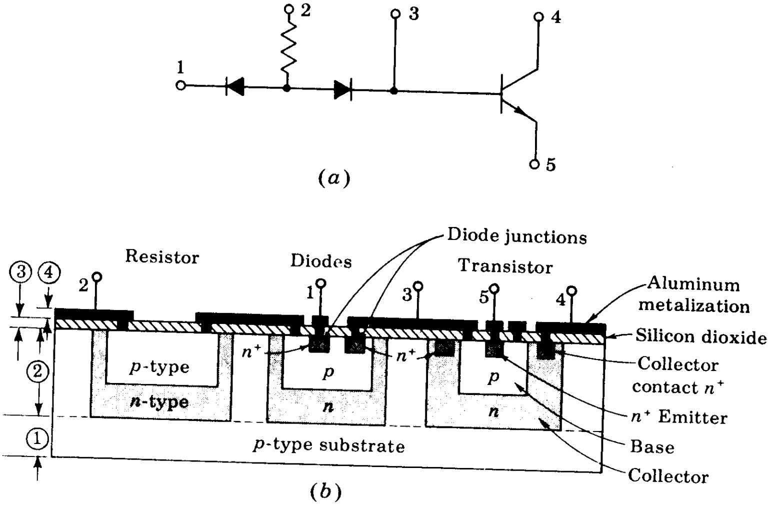

2
3
4
1
5
(a)
Diode junctions
Resistor
Diodes
Transistor
③ ④
2
1
3
5
4
Aluminum metalization
Silicon dioxide
n^+
n^+
p-type
②
n-type
p
n
p
n
Collector contact n^+
n^+ Emitter
①
p-type substrate
Base
Collector
(b)

Fig. 7-1 (*a*) A circuit containing a resistor, two diodes, and a transistor. (*b*) Cross-sectional view of the circuit in (*a*) when transformed into a monolithic form (not drawn to scale). The four layers are ① substrate, ② n-type crystal containing the integrated circuit, ③ silicon dioxide, and ④ aluminum metalization. (After Phillips.[2] Not drawn to scale.)

10 Ω-cm, corresponding to $N_A = 1.4 \times 10^{15}$ atoms/cm³. The epitaxial process described in Sec. 7-3 indicates that the resistivity of the *n*-type epitaxial layer can be chosen independently of that of the substrate. Values of 0.1 to 0.5 Ω-cm are chosen for the *n*-type layer. After polishing and cleaning, a thin layer (0.5 μm = 5,000 Å) of oxide, SiO_2, is formed over the entire wafer, as shown in Fig. 7-2*a*. The SiO_2 is grown by exposing the epitaxial layer to an oxygen atmosphere while being heated to about 1000°C. Silicon dioxide has the fundamental property of preventing the diffusion of impurities through it. Use of this property is made in the following steps.

Step 2. Isolation Diffusion In Fig. 7-2*b* the wafer is shown with the oxide removed in four different places on the surface. This removal is accomplished by means of a photolithographic etching process described in Sec. 7-4. The remaining SiO_2 serves as a mask for the diffusion of acceptor impurities (in this case, boron). The wafer is now subjected to the so-called *isolation diffusion,* which takes place at the temperature and for the time interval required for the *p*-type impurities to penetrate the *n*-type epitaxial layer and reach the *p*-type substrate. We thus leave the shaded *n*-type regions in Fig. 7-2*b*. These sections are called *isolation islands,* or *isolated regions,* because they are separated by two back-to-back *p-n* junctions. Their purpose is to allow electrical isolation between different circuit components. For example, it will become apparent later in this section that a different isolation region must be used for the collector of each separate transistor. The *p*-type subtrate must always be held at a negative potential with respect to the isolation islands in order that the *p-n* junctions be reverse-biased. If these diodes were to become forward-biased in an operating circuit, then, of course, the isolation would be lost.

It should be noted that the concentration of acceptor atoms ($N_A \approx 5 \times 10^{20}$ cm^{-3}) in the region between isolation islands will generally be much higher (and hence indicated as p^+) then in the *p*-type substrate. The reason for this higher density is to prevent the depletion region of the reverse-biased isolation-to-substrate junction from extending into p^+-type material (Sec. 3-7) and possibly connecting two isolation islands.

Parasitic Capacitance It is now important to consider that these isolation regions, or junctions, are connected by a significant barrier, or transition capacitance C_{Ts}, to the *p*-type substrate, which capacitance can affect the operation of the circuit. Since C_{Ts} is an undesirable by-product of the isolation process, it is called the *parasitic capacitance.*

The parasitic capacitance is the sum of two components, the capacitance C_1 from the bottom of the *n*-type region to the substrate (Fig. 7-2*b*) and C_2 from the sidewalls of the isolation islands to the p^+ region. The bottom component, C_1, results from an essentially step junction due to the epitaxial growth (Sec. 7-3), and hence varies inversely as the square root of the voltage V between the isolation region and the substrate (Sec. 3-7). The sidewall capacitance C_2 is associated with a diffused graded junction, and it varies as $V^{-\frac{1}{3}}$.

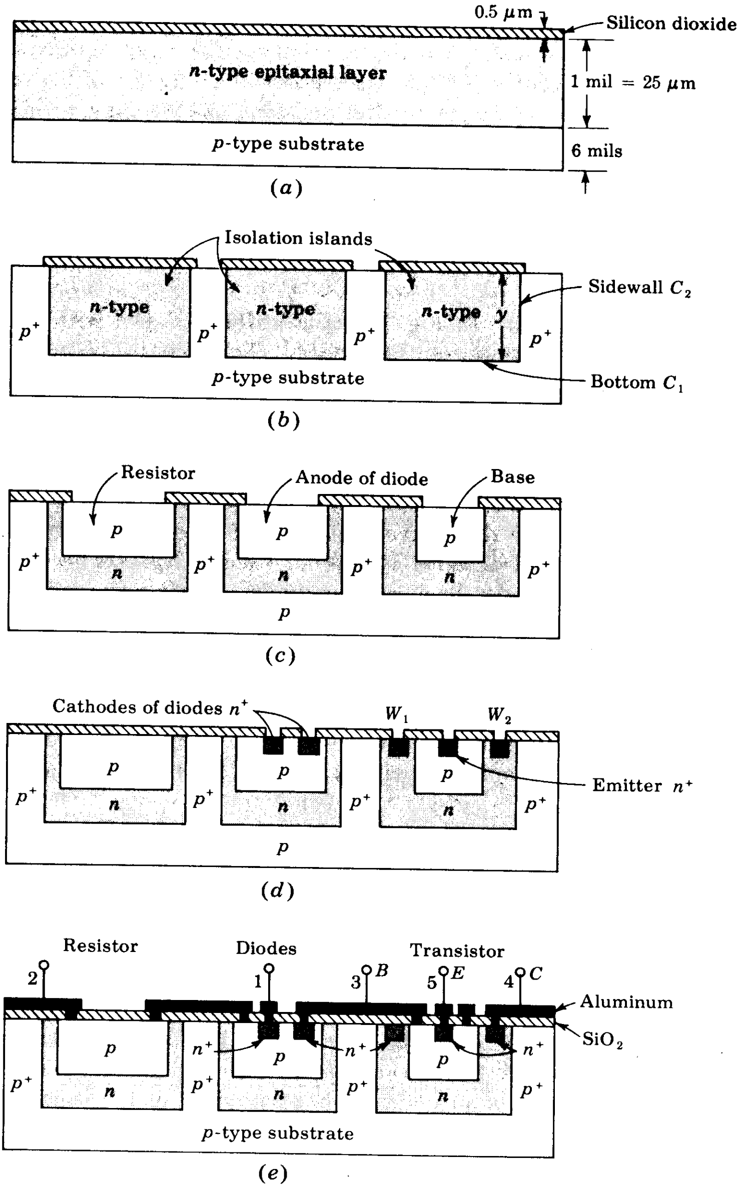

0.5 μm
Silicon dioxide
n-type epitaxial layer
1 mil = 25 μm
p-type substrate
6 mils
(a)
Isolation islands
n-type
n-type
n-type
y
Sidewall C_2
p⁺
p⁺
p⁺
p⁺
p-type substrate
Bottom C_1
(b)
Resistor
Anode of diode
Base
p
p
p
n
n
n
p⁺
p⁺
p⁺
p⁺
p
(c)
Cathodes of diodes n⁺
W_1
W_2
Emitter n⁺
p
p
p
n
n
n
p⁺
p⁺
p⁺
p⁺
p
(d)
Resistor
Diodes
Transistor
2
1
3 B
5 E
4 C
Aluminum
SiO₂
n⁺
n⁺
n⁺
p
p
p
n
n
n
p⁺
p⁺
p⁺
p⁺
p-type substrate
(e)

Fig. 7-2 The steps involved in fabricating a monolithic circuit (not drawn to scale). (a) Epitaxial growth; (b) isolation diffusion; (c) base diffusion; (d) emitter diffusion; (e) aluminum metalization.

For this component the junction area is equal to the perimeter of the isolation region times the thickness y of the epitaxial n-type layer. The total capacitance is of the order of a few picofarads.

Step 3. Base Diffusion During this process a new layer of oxide is formed over the wafer, and the photolithographic process is used again to create the pattern of openings shown in Fig. 7-2*c*. The p-type impurities (boron) are diffused through these openings. In this way are formed the transistor base regions as well as resistors, the anode of diodes, and junction capacitors (if any). It is important to control the depth of this diffusion so that it is shallow and does not penetrate to the substrate. The resistivity of the base layer will generally be much higher than that of the isolation regions.

Step 4. Emitter Diffusion A layer of oxide is again formed over the entire surface, and the masking and etching processes are used again to open windows in the p-type regions, as shown in Fig. 7-2*d*. Through these openings are diffused n-type impurities (phosphorus) for the formation of transistor emitters, the cathode regions for diodes, and junction capacitors.

Additional windows (such as W_1 and W_2 in Fig. 7-2*d*) are often made into the n regions to which a lead is to be connected, using aluminum as the ohmic contact, or interconnecting metal. During the diffusion of phosphorus a heavy concentration (called n^+) is formed at the points where contact with aluminum is to be made. Aluminum is a p-type impurity in silicon, and a large concentration of phosphorus prevents the formation of a p-n junction when the aluminum is alloyed to form an ohmic contact.[4]

Step 5. Aluminum Metalization All p-n junctions and resistors for the circuit of Fig. 7-1*a* have been formed in the preceding steps. It is now necessary to interconnect the various components of the integrated circuit as dictated by the desired circuit. To make these connections, a fourth set of windows is opened into a newly formed SiO_2 layer, as shown in Fig. 7-2*e*, at the points where contact is to be made. The interconnections are made first, using vacuum deposition of a thin even coating of aluminum over the entire wafer. The photoresist technique is now applied to etch away all undesired aluminum areas, leaving the desired pattern of interconnections shown in Fig. 7-2*e* between resistors, diodes, and transistors.

In production a large number (several hundred) of identical circuits such as that of Fig. 7-1*a* are manufactured simultaneously on a single wafer. After the metalization process has been completed, the wafer is scribed with a diamond-tipped tool and separated into individual chips. Each chip is then mounted on a ceramic wafer and is attached to a suitable header. The package leads are connected to the integrated circuit by stitch bonding[1] of a 1-mil aluminum or gold wire from the terminal pad on the circuit to the package lead (Fig. 7-27).

Summary In this section the epitaxial-diffused method of fabricating integrated circuits is described. We have encountered the following processes:

1. Crystal growth of a substrate
2. Epitaxy
3. Silicon dioxide growth
4. Photoetching
5. Diffusion
6. Vacuum evaporation of aluminum

Using these techniques, it is possible to produce the following elements on the same chip: transistors, diodes, resistors, capacitors, and aluminum interconnections.

7-3 EPITAXIAL GROWTH[1]

The epitaxial process produces a thin film of single-crystal silicon from the gas phase upon an existing crystal wafer of the same material. The epitaxial layer may be either p-type or n-type. The growth of an epitaxial film with impurity atoms of boron being trapped in the growing film is shown in Fig. 7-3.

The basic chemical reaction used to describe the epitaxial growth of pure silicon is the hydrogen reduction of silicon tetrachloride:

$$SiCl_4 + 2H_2 \overset{1200°C}{\rightleftarrows} Si + 4HCl \tag{7-1}$$

Since it is required to produce epitaxial films of specific impurity concentrations, it is necessary to introduce impurities such as phosphine for n-type doping or biborane for p-type doping into the silicon tetrachloride–hydrogen

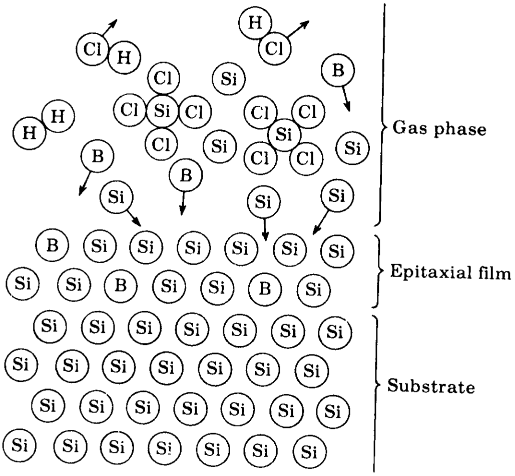


Fig. 7-3 The epitaxial growth of an epitaxial film showing impurity (boron) atoms being trapped in the growing film. (Courtesy of Motorola, Inc.[1])

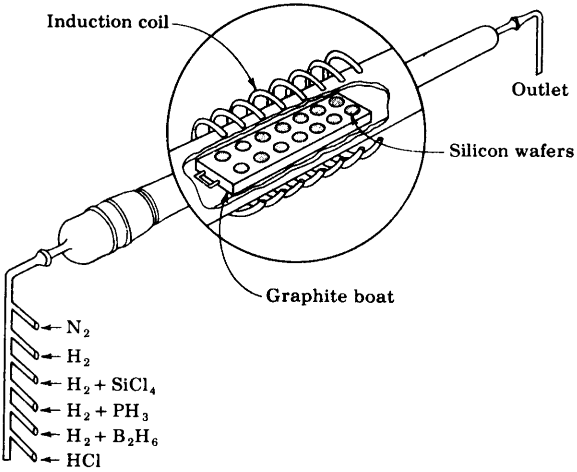


Fig. 7-4 A diagrammatic representation of a system for production growth of silicon epitaxial films. (Courtesy of Motorola, Inc.[1])

gas stream. An apparatus for the production of an epitaxial layer is shown in Fig. 7-4. In this system a long cylindrical quartz tube is encircled by a radio-frequency induction coil. The silicon wafers are placed on a rectangular graphite rod called a *boat*. The boat is inserted in the reaction chamber, and the graphite is heated inductively to about 1200°C. At the input of the reaction chamber a control console permits the introduction of various gases required for the growth of appropriate epitaxial layers. Thus it is possible to form an almost abrupt step *p-n* junction similar to the junction shown in Fig. 3-10.

7-4 MASKING AND ETCHING[1]

The monolithic technique described in Sec. 7-2 requires the selective removal of the SiO_2 to form openings through which impurities may be diffused. The photoetching method used for this removal is illustrated in Fig. 7-5. During the photolithographic process the wafer is coated with a uniform film of a photosensitive emulsion (such as the Kodak *photoresist* KPR). A large black-and-white layout of the desired pattern of openings is made and then reduced photographically. This negative, or stencil, of the required dimensions is placed as a mask over the photoresist, as shown in Fig. 7-5*a*. By exposing the KPR to ultraviolet light through the mask, the photoresist becomes polymerized under the transparent regions of the stencil. The mask is now removed, and the wafer is "developed" by using a chemical (such as trichloroethylene) which dissolves the unexposed (unpolymerized) portions of the photoresist film and leaves the surface pattern as shown in Fig. 7-5*b*. The emulsion which was not removed in development is now *fixed*, or *cured*, so that it becomes resistant to the corrosive etches used next. The chip is

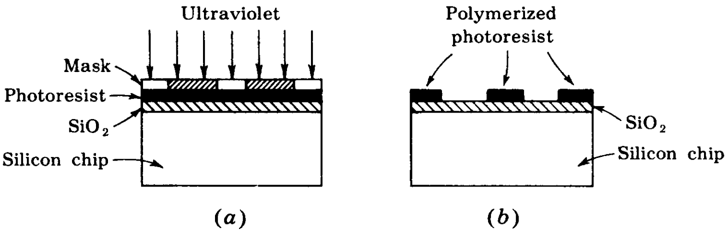


Fig. 7-5 Photoetching technique. (a) Masking and exposure to ultraviolet radiation. (b) The photoresist after development.

immersed in an etching solution of hydrofluoric acid, which removes the oxide from the areas through which dopants are to be diffused. Those portions of the SiO_2 which are protected by the photoresist are unaffected by the acid. After etching and diffusion of impurities, the resist mask is removed (stripped) with a chemical solvent (hot H_2SO_4) and by means of a mechanical abrasion process.

7-5 DIFFUSION OF IMPURITIES[5]

The most important process in the fabrication of integrated circuits is the diffusion of impurities into the silicon chip. We now examine the basic theory connected with this process. The solution to the diffusion equation will give the effect of temperature and time on the diffusion distribution.

The Diffusion Law The equation governing the diffusion of neutral atoms is

$$\frac{\partial N}{\partial t} = D\frac{\partial^2 N}{\partial x^2} \qquad (7\text{-}2)$$

where N is the particle concentration in atoms per unit volume as a function of distance x from the surface and time t, and D is the diffusion constant in area per unit time.

The Complementary Error Function If an intrinsic silicon wafer is exposed to a volume of gas having a uniform concentration N_o atoms per unit volume of n-type impurities, such as phosphorus, these atoms will diffuse into the silicon crystal, and their distribution will be as shown in Fig. 7-6a. If the diffusion is allowed to proceed for extremely long times, the silicon will become uniformly doped with N_o phosphorus atoms per unit volume. The basic assumptions made here are that the surface concentration of impurity atoms remains at N_o for all diffusion times and that $N(x) = 0$ at $t = 0$ for $x > 0$.

If Eq. (7-2) is solved and the above boundary conditions are applied,

$$N(x, t) = N_o\left(1 - \operatorname{erf}\frac{x}{2\sqrt{Dt}}\right) = N_o \operatorname{erfc}\frac{x}{2\sqrt{Dt}} \tag{7-3}$$

where erfc y means the error-function complement of y, and the *error function* of y is defined by

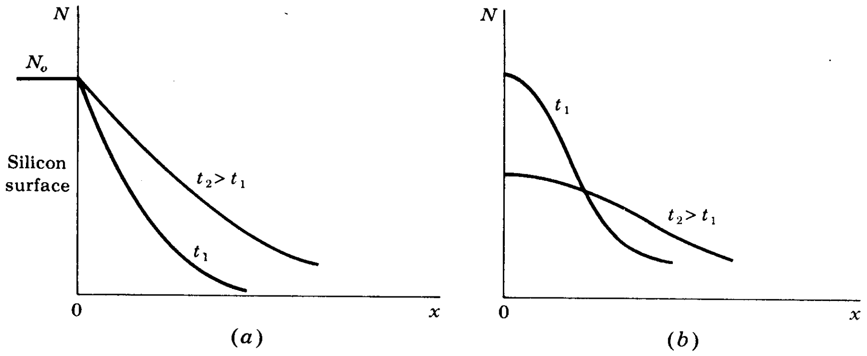
$$\operatorname{erf} y \equiv \frac{2}{\sqrt{x}}\int_0^y \epsilon^{-\lambda^2}\,d\lambda \tag{7-4}$$

and is tabulated in Ref. 3. The function erfc $y = 1 - \operatorname{erf} y$ is plotted in Fig. 7-7.

The Gaussian Distribution If a specific number Q of impurity atoms per unit area are deposited on one face of the wafer, and then if the material is heated, the impurity atoms will again diffuse into the silicon. When the boundary conditions $\int_0^\infty N(x)\,dx = Q$ for all times and $N(x) = 0$ at $t = 0$ for $x > 0$ are applied to Eq. (7-2), we find

$$N(x, t) = \frac{Q}{\sqrt{\pi Dt}}\,\epsilon^{-x^2/4Dt} \tag{7-5}$$

Equation (7-5) is known as the Gaussian distribution, and is plotted in Fig. 7-6*b* for two times. It is noted from the figure that as time increases, the surface concentration decreases. The area under each curve is the same, however, since this area represents the total amount of impurity being diffused, and



(*a*) (*b*)

Fig. 7-6 The concentration N as a function of distance x into a silicon chip for two values t_1 and t_2 of the diffusion time. (a) The surface concentration is held constant at N_o per unit volume. (b) The total number of atoms on the surface is held constant at Q per unit area.

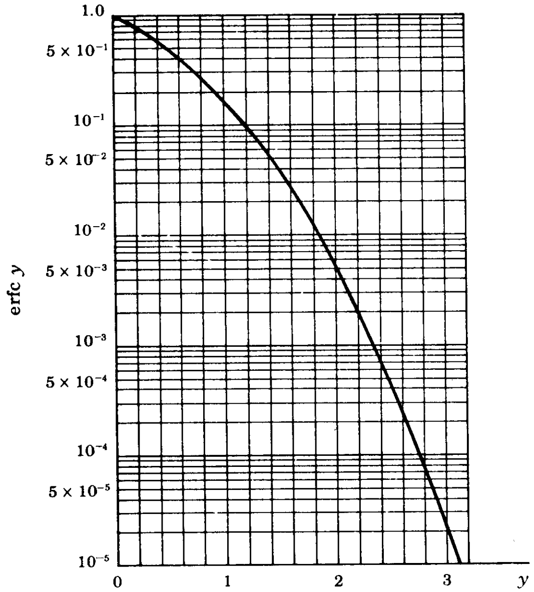


Fig. 7-7 The complementary error function plotted on semilogarithmic paper.

this is a constant amount Q. Note that in Eqs. (7-3) and (7-5) time t and the diffusion constant D appear only as a product Dt.

Solid Solubility[1, 6] The designer of integrated circuits may wish to produce a specific diffusion profile (say, the complementary error function of an n-type impurity). In deciding which of the available impurities (such as phosphorus, arsenic, antimony) can be used, it is necessary to know if the number of atoms per unit volume required by the specific profile of Eq. (7-3) is less than the diffusant's *solid solubility*. The solid solubility is defined as the maximum concentration N_o of the element which can be dissolved in the solid silicon at a given temperature. Figure 7-8 shows solid solubilities of some impurity

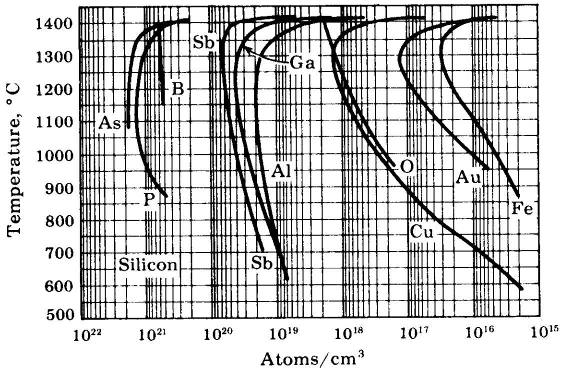


Fig. 7-8 Solid solubilities of some impurity elements in silicon. (After Trumbore,[6] courtesy of Motorola, Inc.[1])

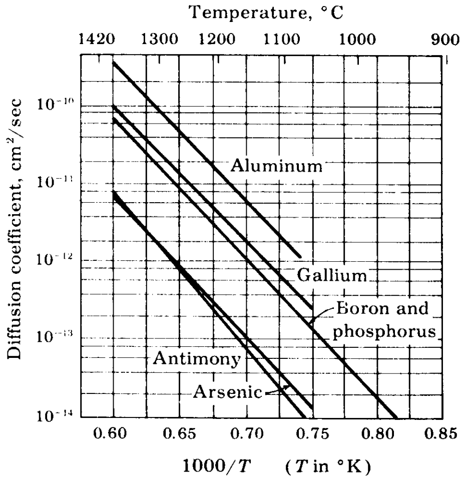


Fig. 7-9 Diffusion coefficients as a function of temperature for some impurity elements in silicon. (After Fuller and Ditzenberger,[5] courtesy of Motorola, Inc.[1])

elements. It can be seen that, since for phosphorus the solid solubility is approximately 10^{21} atoms/cm^3 and for pure silicon we have 5×10^{22} atoms/cm^3, the maximum concentration of phosphorus in silicon is 2 percent. For most of the other impurity elements the solubility is a small fraction of 1 percent.

Diffusion Coefficients Temperature affects the diffusion process because higher temperatures give more energy, and thus higher velocities, to the diffusant atoms. It is clear that the diffusion coefficient is a function of temperature, as shown in Fig. 7-9. From this figure it can be deduced that the diffusion coefficient could be doubled for a few degrees increase in temperature. This critical dependence of D on temperature has forced the development of accurately controlled diffusion furnaces, where temperatures in the range of 1000 to 1300°C can be held to a tolerance of ±0.5°C or better. Since time t in Eqs. (7-3) and (7-5) appears in the product Dt, an increase in either diffusion constant or diffusion time has the same effect on diffusant density.

Note from Fig. 7-9 that the diffusion coefficients, for the same temperature, of the n-type impurities (antimony and arsenic) are lower than the coefficients for the p-type impurities (gallium and aluminum), but that phosphorus (n-type) and boron (p-type) have the same diffusion coefficients.

Typical Diffusion Apparatus Reasonable diffusion times require high diffusion temperatures (~1000°C). Therefore a high-temperature diffusion furnace, having a closely controlled temperature over the length (20 in.) of the hot zone of the furnace, is standard equipment in a facility for the fabrication of integrated circuits. Impurity sources used in connection with diffu-

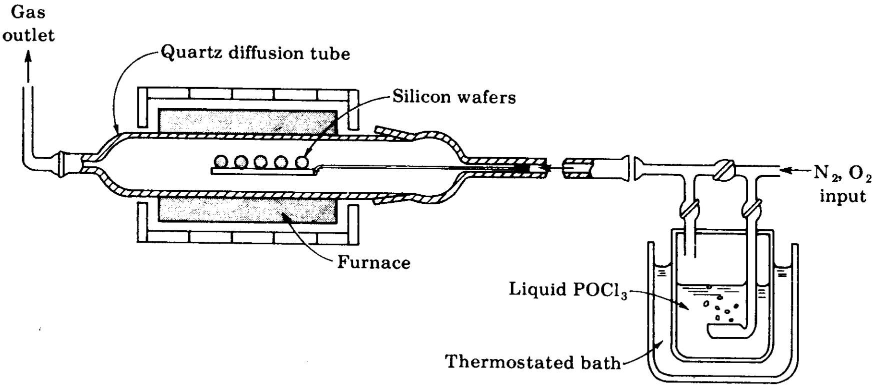


Fig. 7-10 Schematic representation of typical apparatus for $POCl_3$ diffusion. (Courtesy of Motorola, Inc.[1])

sion furnaces can be gases, liquids, or solids. For example, $POCl_3$, which is a liquid, is often used as a source of phosphorus. Figure 7-10 shows the apparatus used for $POCl_3$ diffusion. In this apparatus a carrier gas (mixture of nitrogen and oxygen) bubbles through the liquid-diffusant source and carries the diffusant atoms to the silicon wafers. Using this process, we obtain the complementary-error-function distribution of Eq. (7-3). A two-step procedure is used to obtain the Gaussian distribution. The first step involves *predeposition*, carried out at about 900°C, followed by *drive-in* at about 1100°C.

EXAMPLE A uniformly doped n-type silicon epitaxial layer of 0.5 Ω-cm resistivity is subjected to a boron diffusion with constant surface concentration of 5×10^{18} cm^{-3}. It is desired to form a p-n junction at a depth of 2.7 μm. At what temperature should this diffusion be carried out if it is to be completed in 2 hr?

Solution The concentration N of boron is high at the surface and falls off with distance into the silicon, as indicated in Fig. 7-6*a*. At that distance $x = x_j$ at which N equals the concentration n of the doped silicon wafer, the net impurity density is zero. For $x < x_j$, the net impurity density is positive, and for $x > x_j$, it is negative. Hence x_j represents the distance from the surface at which a junction is formed. We first find n from Eq. (2-8):

$$n = \frac{\sigma}{\mu_n q} = \frac{1}{(0.5)(1{,}300)(1.60 \times 10^{-19})} = 0.96 \times 10^{16}\ \text{cm}^{-3}$$

where all distances are expressed in centimeters and the mobility μ_n for silicon is taken from Table 2-1, on page 29. The junction is formed when $N = n$. For

$$\operatorname{erfc} y = \frac{N}{N_o} = \frac{n}{N_o} = \frac{0.96 \times 10^{16}}{5 \times 10^{18}} = 1.98 \times 10^{-3}$$

we find from Fig. 7-7 that $y = 2.2$. Hence

$$2.2 = \frac{x_j}{2\sqrt{Dt}} = \frac{2.7 \times 10^{-4}}{2\sqrt{D \times 2 \times 3{,}600}}$$

Solving for D, we obtain $D = 5.2 \times 10^{-13}$ cm²/sec. This value of diffusion constant for boron is obtained from Fig. 7-9 at $T = 1130°\text{C}$.

7-6 TRANSISTORS FOR MONOLITHIC CIRCUITS[1,7]

A planar transistor made for monolithic integrated circuits, using epitaxy and diffusion, is shown in Fig. 7-11*a*. Here the collector is electrically separated from the substrate by the reverse-biased isolation diodes. Since the anode of the isolation diode covers the back of the entire wafer, it is necessary to make the collector contact on the top, as shown in Fig. 7-11*a*. It is now clear that the isolation diode of the integrated transistor has two undesirable effects: it adds a parasitic shunt capacitance to the collector and a leakage current path. In addition, the necessity for a top connection for the collector increases the collector-current path and thus increases the collector resistance and $V_{CE,\text{sat}}$. All these undesirable effects are absent from the discrete epitaxial transistor shown in Fig. 7-11*b*. What is then the advantage of the monolithic transistor? A significant improvement in performance arises from the fact that integrated transistors are located physically close together and their electrical characteristics are closely matched. For example, integrated transistors spaced within 30 mils (0.03 in.) have V_{BE} matching of better than

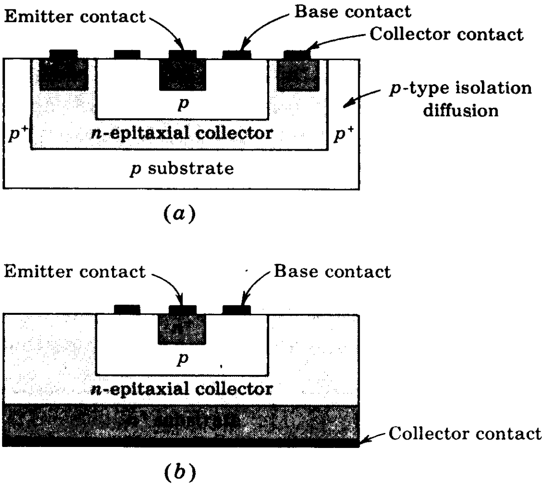

Emitter contact
Base contact
Collector contact
p-type isolation diffusion
p
p+
n-epitaxial collector
p+
p substrate
(a)
Emitter contact
Base contact
p
n-epitaxial collector
n+ substrate
Collector contact
(b)

Fig. 7-11 Comparison of cross sections of (*a*) a monolithic integrated circuit transistor with (*b*) a discrete planar epitaxial transistor. [For a top view of the transistor in (*a*) see Fig. 7-13.]

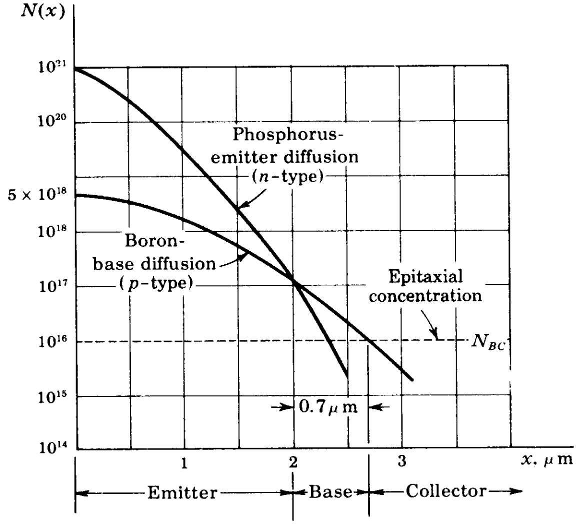


Fig. 7-12 A typical impurity profile in a monolithic integrated transistor. [Note that $N(x)$ is plotted on a logarithmic scale.]

5 mV with less than 10 μV/°C drift and an h_{FE} match of ±10 percent. These matched transistors make excellent difference amplifiers (Sec. 15-3).

The electrical characteristics of a transistor depend on the size and geometry of the transistor, doping levels, diffusion schedules, and the basic silicon material. Of all these factors the size and geometry offer the greatest flexibility for design. The doping levels and diffusion schedules are determined by the standard processing schedule used for the desired transistors in the integrated circuit.

Impurity Profiles for Integrated Transistors[1] Figure 7-12 shows a typical impurity profile for a monolithic integrated circuit transistor. The background, or epitaxial-collector, concentration N_{BC} is shown as a dashed line in Fig. 7-12. The base diffusion of p-type impurities (boron) starts with a surface concentration of 5×10^{18} atoms/cm^3, and is diffused to a depth of 2.7 μm, where the collector junction is formed. The emitter diffusion (phosphorus) starts from a much higher surface concentration (close to the solid solubility) of about 10^{21} atoms/cm^3, and is diffused to a depth of 2 μm, where the emitter junction is formed. This junction corresponds to the intersection of the base and emitter distribution of impurities. We now see that the base thickness for this monolithic transistor is 0.7 μm. The emitter-to-base junction is usually treated as a step-graded junction, whereas the base-to-collector junction is considered a linearly graded junction.

EXAMPLE (*a*) Obtain the equations for the impurity profiles in Fig. 7-12. (*b*) If the phosphorus diffusion is conducted at 1100°C, how long should be allowed for this diffusion?

Solution *a.* The base diffusion specifications are exactly those given in the example on page 208, where we find (with x expressed in micrometers) that

$$y = 2.2 = \frac{2.7}{2\sqrt{Dt}}$$

or

$$2\sqrt{Dt} = \frac{2.7}{2.2} = 1.23\ \mu\text{m}$$

Hence the boron profile, given by Eq. (7-3), is

$$N_B = 5 \times 10^{18} \operatorname{erfc} \frac{x}{1.23}$$

The emitter junction is formed at $x = 2\ \mu$m, and the boron concentration here is

$$N_B = 5 \times 10^{18} \operatorname{erfc} \frac{2}{1.23} = 5 \times 10^{18} \times 2 \times 10^{-2}$$
$$= 1.0 \times 10^{17}\ \text{cm}^{-3}$$

The phosphorus concentration N_P is given by

$$N_P = 10^{21} \operatorname{erfc} \frac{x}{2\sqrt{Dt}}$$

At $x = 2$, $N_P = N_B = 1.0 \times 10^{17}$, so that

$$\operatorname{erfc} \frac{2}{2\sqrt{Dt}} = \frac{1.0 \times 10^{17}}{10^{21}} = 1.0 \times 10^{-4}$$

From Fig. 7-7, $2/(2\sqrt{Dt}) = 2.7$ and $2\sqrt{Dt} = 0.75\ \mu$m. Hence the phosphorus profile is given by

$$N_P = 10^{21} \operatorname{erfc} \frac{x}{0.74}$$

b. From Fig. 7-9, at $T = 1100°\text{C}$, $D = 3.8 \times 10^{-13}$ cm^2/sec. Solving for t from $2\sqrt{Dt} = 0.74\ \mu$m, we obtain

$$t = \frac{(0.37 \times 10^{-4})^2}{3.8 \times 10^{-13}} = 3{,}600\ \text{s} = 60\ \text{min}$$

Monolithic Transistor Layout[1,2] The physical size of a transistor determines the parasitic isolation capacitance as well as the junction capacitance. It is therefore necessary to use small-geometry transistors if the integrated circuit is designed to operate at high frequencies or high switching speeds. The geometry of a typical monolithic transistor is shown in Fig. 7-13. The emitter rectangle measures 1 by 1.5 mils, and is diffused into a 2.5- by 4.0-mil base region. Contact to the base is made through two metalized stripes on either side of the emitter. The rectangular metalized area forms the ohmic

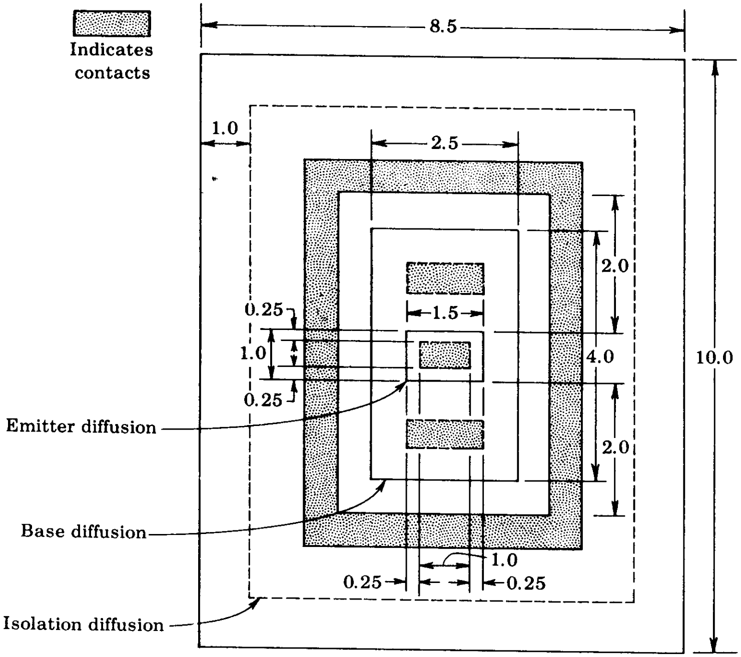


Fig. 7-13 A typical double-base stripe geometry of an integrated-circuit transistor. Dimensions are in mils. (For a side view of the transistor see Fig. 7-11.) (Courtesy of Motorola Monitor.)

contact to the collector region. The rectangular collector contact of this transistor reduces the saturation resistance. The substrate in this structure is located about 1 mil below the surface. Since diffusion proceeds in three dimensions, it is clear that the *lateral-diffusion* distance will also be 1 mil. The dashed rectangle in Fig. 7-13 represents the substrate area and is 6.5 by 8 mils. A summary of the electrical properties[2] of this transistor for both the 0.5- and the 0.1-Ω-cm collectors is given in Table 7-1.

Buried Layer[1] We noted above that the integrated transistor, because of the top collector contact, has a higher collector series resistance than a similar discrete-type transistor. One common method of reducing the collector series resistance is by means of a heavily doped n^+ "buried" layer sandwiched between the p-type substrate and the n-type epitaxial collector, as shown in Fig. 7-14. The buried-layer structure can be obtained by diffusing the n^+ layer into the substrate before the n-type epitaxial collector is grown or by selectively growing the n^+-type layer, using masked epitaxial techniques.

We are now in a position to appreciate one of the reasons why the integrated transistor is usually of the n-p-n type. Since the collector region is

TABLE 7-1 Characteristics for 1- by 1.5-mil double-base stripe monolithic transistors[2]

Transistor parameter	0.5 Ω-cm	0.1 Ω-cm†
BV_{CBO}, V	55	25
BV_{EBO}, V	7	5.5
BV_{CEO}, V	23	14
$C_{Te,\text{forward bias}}$, pF	6	10
C_{Te} at 0.5 V, pF	1.5	2.5
C_{Te} at 5 V, pF	0.7	1.5
h_{FE} at 10 mA	50	50
R_{CS}, Ω	75	15
$V_{CE,\text{sat}}$ at 5 mA, V	0.5	0.26
V_{BE} at 10 mA, V	0.85	0.85
f_T at 5 V, 5 mA, MHz	440	520

† Gold-doped.

subjected to heating during the base and emitter diffusions, it is necessary that the diffusion coefficient of the collector impurities be as small as possible, to avoid movement of the collector junction. Since Fig. 7-9 shows that n-type impurities have smaller values of diffusion constant D than p-type impurities, the collector is usually n-type. In addition, the solid solubility of some n-type impurities is higher than that of any p-type impurity, thus allowing heavier doping of the n^+-type emitter and other n^+ regions.

Lateral p-n-p Transistor[9] The standard integrated-circuit transistor is an n-p-n type, as we have already emphasized. In some applications it is required to have both n-p-n and p-n-p transistors on the same chip. The lateral p-n-p structure shown in Fig. 7-15 is the most common form of the integrated p-n-p transistor. This p-n-p uses the standard diffusion techniques as the n-p-n, but the last n diffusion (used for the n-p-n transistor) is eliminated. While the p base for the n-p-n transistor is made, the two adjacent p regions are diffused for the emitter and collector of the p-n-p transistor shown in Fig. 7-15. Note that the current flows *laterally* from emitter to collector. Because of inaccuracies in masking, and because, also, of lateral diffusion, the base width between emitter and collector is large (about 1 mil compared with 1 μm

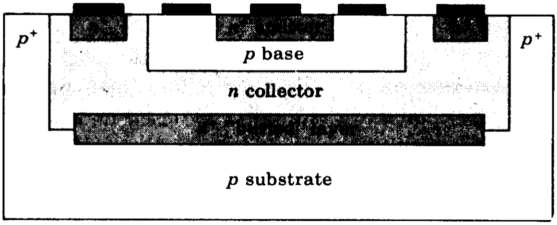


Fig. 7-14 Utilization of "buried" n^+ layer to reduce collector series resistance.

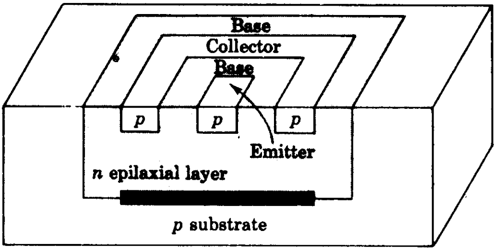


Fig. 7-15 A p-n-p lateral transistor.

for an n-p-n base). Hence the current gain of the p-n-p transistor is very low (0.5 to 5) instead of 50 to 300 for the n-p-n device. Since the base-p resistivity of the n-p-n transistor is relatively high, the collector and emitter resistances of the p-n-p device are high.

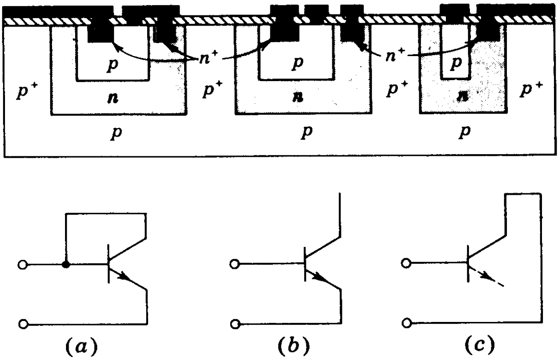
Vertical p-n-p Transistor[9] This transistor uses the substrate for the p collector; the n epitaxial layer for the base; and the p base of the standard n-p-n transistor as the emitter of this p-n-p device. We have already emphasized that the substrate must be connected to the most negative potential in the circuit. Hence a vertical p-n-p transistor can be used only if its collector is at a fixed negative voltage. Such a configuration is called an *emitter follower*, and is discussed in Sec. 8-8.

Supergain n-p-n Transistor[9] If the emitter is diffused into the base region so as to reduce the effective base width almost to the point of *punch-through* (Sec. 5-13), the current gain may be increased drastically (typically, 5,000). However, the breakdown voltage is reduced to a very low value (say, 5 V). If such a transistor in the CE configuration is operated in series with a standard integrated CB transistor (such a combination is called a *cascode* arrangement), the superhigh gain can be obtained at very low currents and with breakdown voltages in excess of 50 V.

7-7 MONOLITHIC DIODES[1]

The diodes utilized in integrated circuits are made by using transistor structures in one of five possible connections (Prob. 7-9). The three most popular diode structures are shown in Fig. 7-16. They are obtained from a transistor structure by using the emitter-base diode, with the collector short-circuited to the base (*a*); the emitter-base diode, with the collector open (*b*); and the collector-base diode, with the emitter open-circuited (or not fabricated at all) (*c*). The choice of the diode type used depends upon the application and circuit performance desired. Collector-base diodes have the higher collector-base voltage-breaking rating of the collector junction (~12 V minimum), and they are suitable for common-cathode diode arrays diffused within a single isolation

Fig. 7-16 Cross section of various diode structures. (a) Emitter-base diode with collector shorted to base; (b) emitter-base diode with collector open; (c) collector-base diode (no emitter diffusion).



island, as shown in Fig. 7-17*a*. Common-anode arrays can also be made with the collector-base diffusion, as shown in Fig. 7-17*b*. A separate isolation is required for each diode, and the anodes are connected by metalization.

The emitter-base diffusion is very popular for the fabrication of diodes provided that the reverse-voltage requirement of the circuit does not exceed the lower base-emitter breakdown voltage (~7 V). Common-anode arrays can easily be made with the emitter-base diffusion by using a multiple-emitter transistor within a single isolation area, as shown in Fig. 7-18. The collector may be either open or shorted to the base. The diode pair in Fig. 7-1 is constructed in this manner, with the collector floating (open).

Diode Characteristics The forward volt-ampere characteristics of the three diode types discussed above are shown in Fig. 7-19. It will be observed that the diode-connected transistor (emitter-base diode with collector shorted

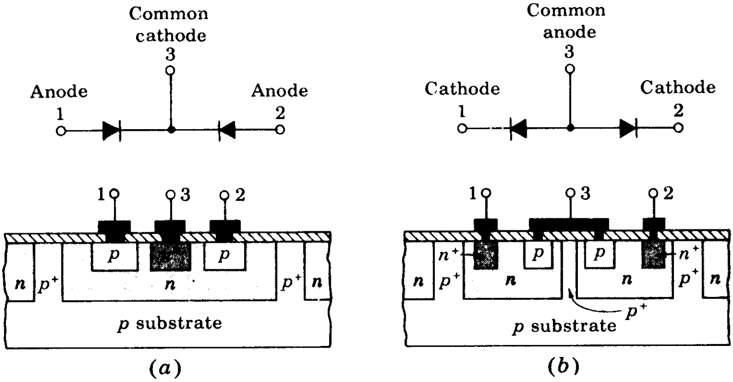


Fig. 7-17 Diode pairs. (a) Common-cathode pair and (b) common-anode pair, using collector-base diodes.

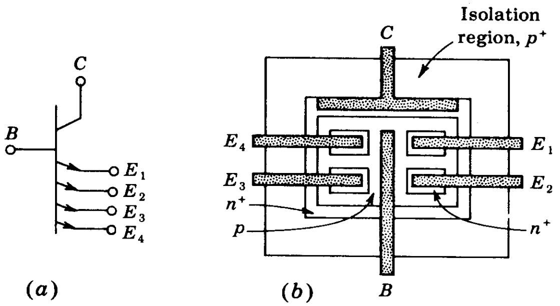


Fig. 7-18 A multiple-emitter *n-p-n* transistor. (*a*) Schematic, (*b*) monolithic surface pattern. If the base is connected to the collector, the result is a multiple-cathode diode structure with a common anode.

to the base) provides the highest conduction for a given forward voltage. The reverse recovery time for this diode is also smaller, one-third to one-fourth that of the collector-base diode.

7-8 INTEGRATED RESISTORS[1]

A resistor in a monolithic integrated circuit is very often obtained by utilizing the bulk resistivity of one of the diffused areas. The *p*-type base diffusion is most commonly used, although the *n*-type emitter diffusion is also employed. Since these diffusion layers are very thin, it is convenient to define a quantity known as the *sheet resistance* R_S.

Sheet Resistance If, in Fig. 7-20, the width w equals the length l, we have a square l by l of material with resistivity ρ, thickness y, and cross-sectional area $A = ly$. The resistance of this conductor (in ohms per square) is

$$R_S = \frac{\rho l}{ly} = \frac{\rho}{y} \tag{7-6}$$

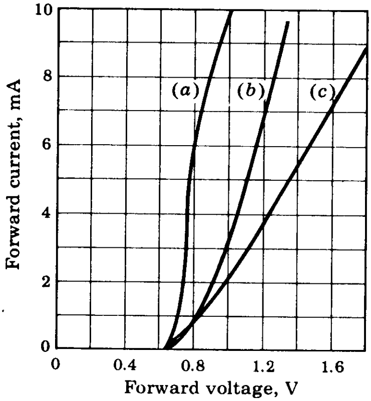


Fig. 7-19 Typical diode volt-ampere characteristics for the three diode types of Fig. 7-16. (*a*) Base-emitter (collector shorted to base); (*b*) base-emitter (collector open); (*c*) collector-base (emitter open). (Courtesy of Fairchild Semiconductor.)

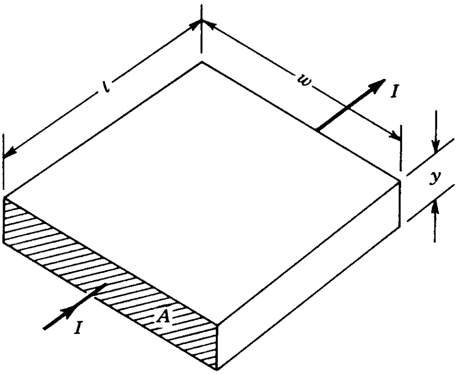


Fig. 7-20 Pertaining to sheet resistance, ohms per square.

Note that R_S is independent of the size of the square. Typically, the sheet resistance of the base and emitter diffusions whose profiles are given in Fig. 7-12 is 200 Ω/square and 2.2 Ω/square, respectively.

The construction of a base-diffused resistor is shown in Fig. 7-1 and is repeated in Fig. 7-21*a*. A top view of this resistor is shown in Fig. 7-21*b*. The resistance value may be computed from

$$R = \frac{\rho l}{yw} = R_S \frac{l}{w} \tag{7-7}$$

where l and w are the length and width of the diffused area, as shown in the top view. For example, a base-diffused-resistor stripe 1 mil wide and 10 mils long contains 10 (1 by 1 mil) squares, and its value is 10 × 200 = 2,000 Ω. Empirical[1,2] corrections for the end contacts are usually included in calculations of R.

Resistance Values Since the sheet resistance of the base and emitter diffusions is fixed, the only variables available for diffused-resistor design are

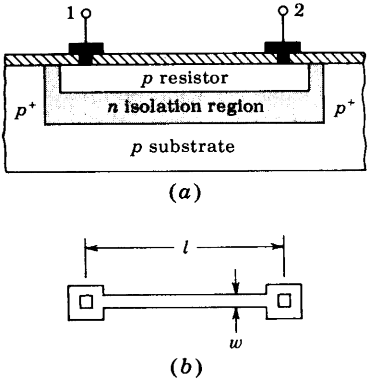


Fig. 7-21 A monolithic resistor. (*a*) Cross-sectional view; (*b*) top view.

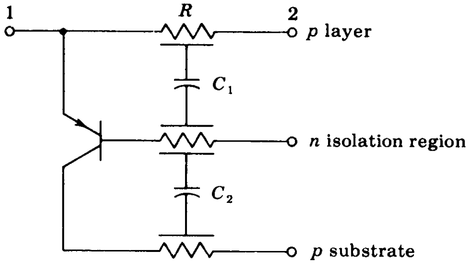


Fig. 7-22 The equivalent circuit of a diffused resistor.

stripe length and stripe width. Stripe widths of less than 1 mil (0.001 in.) are not normally used because a line-width variation of 0.0001 in. due to mask drawing error or mask misalignment or photographic-resolution error can result in 10 percent resistor-tolerance error.

The range of values obtainable with diffused resistors is limited by the size of the area required by the resistor. Practical range of resistance is 20 Ω to 30 K for a base-diffused resistor and 10 Ω to 1 K for an emitter-diffused resistor. The tolerance which results from profile variations and surface geometry errors[1] is as high as ±10 percent of the nominal value at 25°C, with ratio tolerance of ±1 percent. For this reason the design of integrated circuits should, if possible, emphasize resistance ratios rather than absolute values. The temperature coefficient for these heavily doped resistors is positive (for the same reason that gives a positive coefficient to the silicon sensistor, discussed in Sec. 2-7) and is +0.06 percent/°C from −55 to 0°C and +0.20 percent/°C from 0 to 125°C.

Equivalent Circuit A model of the diffused resistor is shown in Fig. 7-22, where the parasitic capacitances of the base-isolation (C_1) and isolation-substrate (C_2) junctions are included. In addition, it can be seen that a parasitic *p-n-p* transistor exists, with the substrate as collector, the isolation *n*-type region as base, and the resistor *p*-type material as the emitter. The collector is reverse-biased because the *p*-type substrate is at the most negative potential. It is also necessary that the emitter be reverse-biased to keep the parasitic transistor at cutoff. This condition is maintained by placing all resistors in the same isolation region and connecting the *n*-type isolation region surrounding the resistors to the *most positive* voltage present in the circuit. Typical values of h_{fe} for this parasitic transistor range from 0.5 to 5.

Thin-film Resistors[1] A technique of vapor thin-film deposition can also be used to fabricate resistors for integrated circuits. The metal (usually nichrome NiCr) film is deposited (to a thickness at less than 1 μm) on the silicon dioxide layer, and masked etching is used to produce the desired geometry. The metal resistor is then covered by an insulating layer, and apertures for the ohmic contacts are opened through this insulating layer. Typical sheet-resistance values for nichrome thin-film resistors are 40 to 400 Ω/square, resulting in resistance values from about 20 Ω to 50 K.

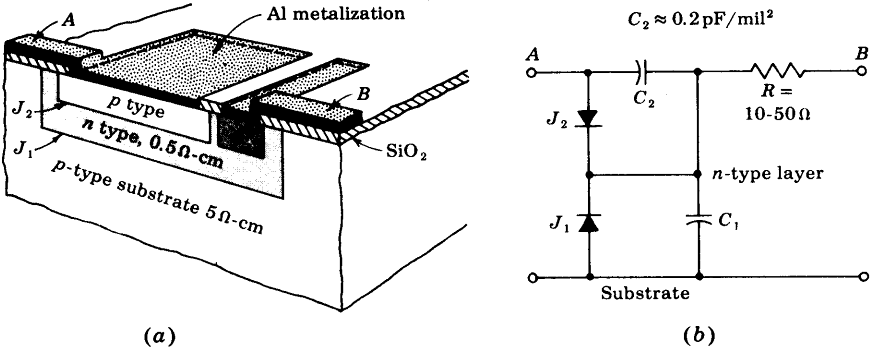
7-9 INTEGRATED CAPACITORS AND INDUCTORS[1,2]

Capacitors in integrated circuits may be obtained by utilizing the transition capacitance of a reverse-biased *p-n* junction or by a thin-film technique.

Junction Capacitors A cross-sectional view of a junction capacitor is shown in Fig. 7-23*a*. The capacitor is formed by the reverse-biased junction J_2, which separates the epitaxial *n*-type layer from the upper *p*-type diffusion area. An additional junction J_1 appears between the *n*-type epitaxial plane and the substrate, and a parasitic capacitance C_1 is associated with this reverse-biased junction. The equivalent circuit of the junction capacitor is shown in Fig. 7-23*b*, where the desired capacitance C_2 should be as large as possible relative to C_1. The value of C_2 depends on the junction area and impurity concentration. Since this junction is essentially abrupt, C_2 is given by Eq. (3-23). The series resistance R (10 to 50 Ω) represents the resistance of the *n*-type layer.

It is clear that the substrate must be at the most negative voltage so as to minimize C_1 and isolate the capacitor from other elements by keeping junction J_1 reverse-biased. It should also be pointed out that the junction capacitor C_2 is polarized since the *p-n* junction J_2 must always be reverse-biased.

Thin-film Capacitors A metal-oxide-semiconductor (MOS) nonpolarized capacitor is indicated in Fig. 7-24*a*. This structure is a parallel-plate capac-



(*a*) (*b*)

Fig. 7-23 (*a*) Junction monolithic capacitor. (*b*) Equivalent circuit. (Courtesy of Motorola, Inc.)

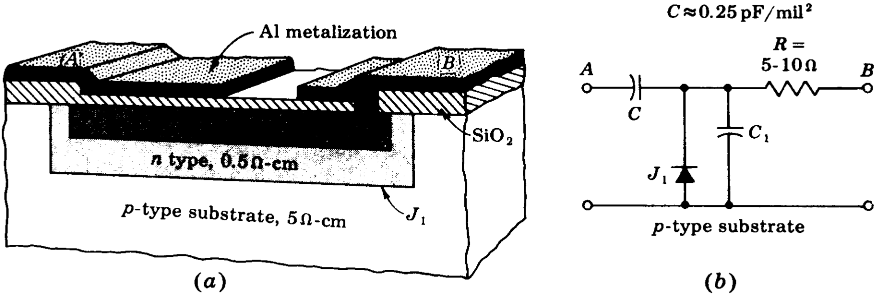


Fig. 7-24 An MOS capacitor. (a) The structure; (b) the equivalent circuit.

itor with SiO_2 as the dielectric. A surface thin film of metal (aluminum) is the top plate. The bottom plate consists of the heavily doped n^+ region that is formed during the emitter diffusion. A typical value for capacitance[8] is 0.4 pF/mil² for an oxide thickness of 500 Å, and the capacitance varies inversely with the thickness.

The equivalent circuit of the MOS capacitor is shown in Fig. 7-24*b*, where C_1 denotes the parasitic capacitance J_1 of the collector-substrate junction, and R is the small series resistance of the n^+ region. Table 7-2 lists the range of possible values for the parameters of junction and MOS capacitors.

TABLE 7-2 Integrated capacitor parameters

Characteristic	Diffused-junction capacitor	Thin-film MOS
Capacitance, pF/mil²	0.2	0.25–0.4
Maximum area, mil²	2×10^3	2×10^3
Maximum value, pF	400	800
Breakdown voltage, V	5–20	50–200
Voltage dependence	$kV^{-\frac{1}{2}}$	0
Tolerance, percent	±20	±20

Inductors No practical inductance values have been obtained at the present time (1972) on silicon substrates using semiconductor or thin-film techniques. Therefore their use is avoided in circuit design wherever possible. If an inductor is required, a discrete component is connected externally to the integrated circuit.

Characteristics of Integrated Components Based upon our discussion of integrated-circuit technology, we can summarize the significant characteristics of integrated circuits (in addition to the advantages listed in Sec. 7-1).

1. A restricted range of values exists for resistors and capacitors. Typically, $10\ \Omega \leq R \leq 30$ K and $C \leq 200$ pF.

2. Poor tolerances are obtained in fabricating resistors and capacitors of specific magnitudes. For example, ±20 percent of absolute values is typical. Resistance ratio tolerance can be specified to ±1 percent because all resistors are made at the same time using the same techniques.

3. Components have high-temperature coefficients and may also be voltage-sensitive.

4. High-frequency response is limited by parasitic capacitances.

5. The technology is very costly for small-quantity production.

6. No practical inductors or transformers can be integrated.

In the next section we examine some of the design rules for the layout of monolithic circuits.

7-10 MONOLITHIC–CIRCUIT LAYOUT[1,10]

In this section we describe how to transform the discrete logic circuit of Fig. 7-25*a* into the layout of the monolithic circuit shown in Fig. 7-26.

Design Rules for Monolithic Layout The following 10 reasonable design rules are stated by Phillips[10]:

1. Redraw the schematic to satisfy the required pin connection with the minimum number of crossovers.

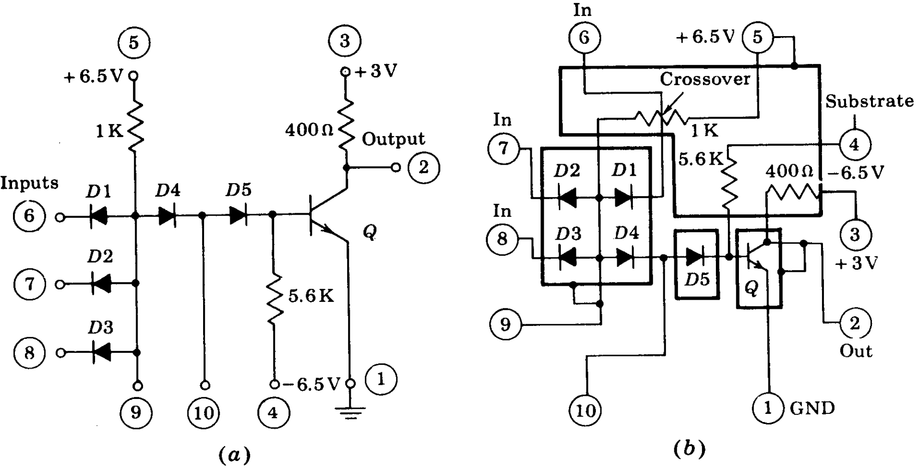


Fig. 7-25 (*a*) A DTL gate. (*b*) The schematic redrawn to indicate the 10 external connections arranged in the sequence in which they will be brought out to the header pins. The isolation regions are shown in heavy outline.

2. Determine the number of isolation islands from collector-potential considerations, and reduce the areas as much as possible.
3. Place all resistors having fixed potentials at one end in the same isolation island, and return that isolation island to the most positive potential in the circuit.
4. Connect the substrate to the most negative potential of the circuit.
5. In layout, allow an isolation border equal to twice the epitaxial thickness to allow for underdiffusion.
6. Use 1-mil widths for diffused emitter regions and $\frac{1}{2}$-mil widths for base contacts and spacings, and for collector contacts and spacings.
7. For resistors, use widest possible designs consistent with die-size limitations. Resistances which must have a close ratio must have the same width and be placed close to one another.
8. Always optimize the layout arrangement to maintain the smallest possible die size, and if necessary, compromise pin connections to achieve this.
9. Determine component geometries from the performance requirements of the circuit.
10. Keep all metalizing runs as short and as wide as possible, particularly at the emitter and collector output connections of the saturating transistor.

Pin Connections The circuit of Fig. 7-25*a* is redrawn in Fig. 7-25*b*, with the external leads labeled 1, 2, 3, . . . , 10 and arranged in the order in which they are connected to the header pins. The diagram reveals that the power-supply pins are grouped together, and also that the inputs are on adjacent pins. In general, the external connections are determined by the system in which the circuits are used.

Crossovers Very often the layout of a monolithic circuit requires two conducting paths (such as leads 5 and 6 in Fig. 7-25*b*) to cross over each other. This crossover cannot be made directly because it will result in electric contact between two parts of the circuit. Since all resistors are protected by the SiO_2 layer, any resistor may be used as a crossover region. In other words, if aluminum metalization is run over a resistor, no electric contact will take place between the resistor and the aluminum.

Sometimes the layout is so complex that additional crossover points may be required. A diffused structure which allows a crossover is also possible.[1]

Isolation Islands The number of isolation islands is determined next. Since the transistor collector requires one isolation region, the heavy rectangle has been drawn in Fig. 7-25*b* around the transistor. It is shown connected to the output pin 2 because this isolation island also forms the transistor collector. Next, all resistors are placed in the same isolation island, and the island is then connected to the most positive voltage in the circuit, for reasons discussed in Sec. 7-8.

To determine the number of isolation regions required for the diodes, it is necessary first to establish which kind of diode will be fabricated. In this case, because of the low forward drop shown in Fig. 7-19, it was decided to make the common-anode diodes of the emitter-base type with the collector shorted to the base. Since the "collector" is at the "base" potential, it is required to have a single isolation island for the four common-anode diodes. Finally, the remaining diode is fabricated as an emitter-base diode, with the collector open-circuited, and thus it requires a separate isolation island.

The Fabrication Sequence The final monolithic layout is determined by a trial-and-error process, having as its objective the smallest possible die size. This layout is shown in Fig. 7-26. The reader should identify the four isolation islands, the three resistors, the five diodes, and the transistor. It is interesting to note that the 5.6-K resistor has been achieved with a 2-mil-wide 1.8-K resistor in series with a 1-mil-wide 3.8-K resistor. To conserve space,

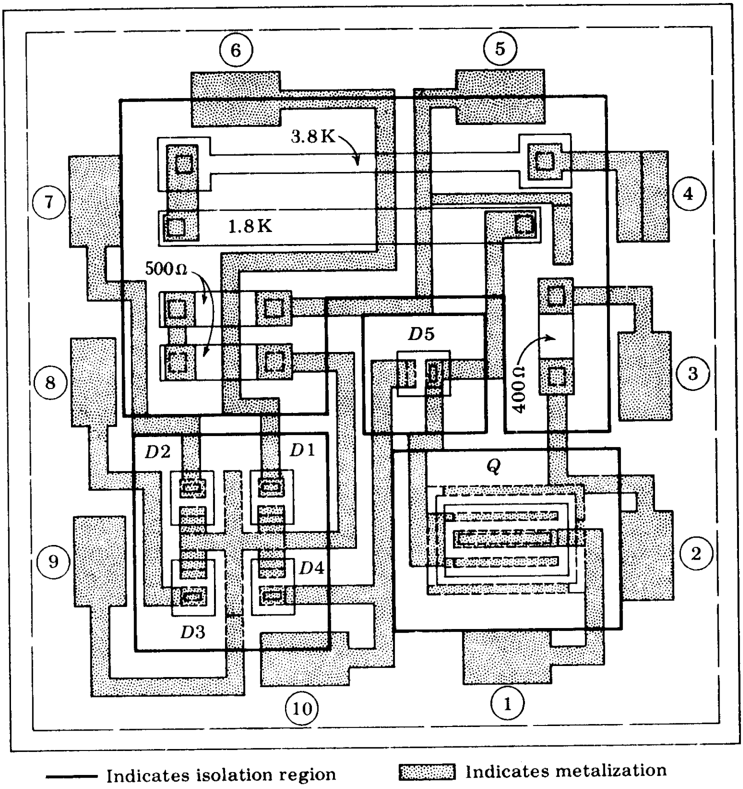
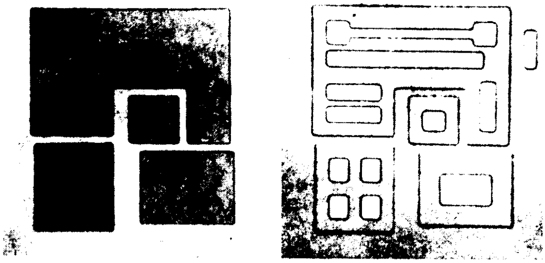
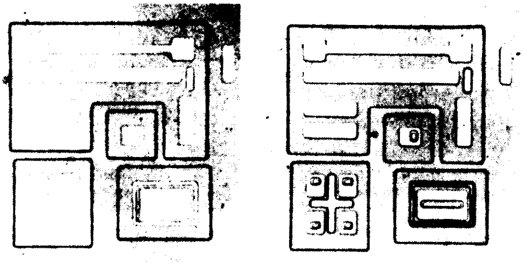


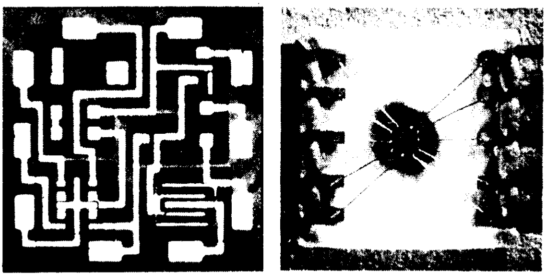
Fig. 7-26 Monolithic design layout for the circuit of Fig. 7-25. (Courtesy of Motorola Monitor, Phoenix, Ariz.)



Isolation diffusion Base diffusion



Emitter diffusion Preohmic etch



Metalization Flat package assembly

Fig. 7-27 Monolithic fabrication sequence for the circuit of Fig. 7-25. (Courtesy of Motorola Monitor, Phoenix, Ariz.)

the resistor was folded back on itself. In addition, two metalizing crossovers ran over this resistor.

From a layout such as shown in Fig. 7-26, the manufacturer produces the masks required for the fabrication of the monolithic integrated circuit. The production sequence which involves isolation, base, and emitter diffusions, preohmic etch, aluminum metalization, and the flat package assembly is shown in Fig. 7-27.

7-11 ADDITIONAL ISOLATION METHODS

Electrical isolation between the different elements of a monolithic integrated circuit is accomplished by means of a diffusion which yields back-to-back *p-n* junctions, as indicated in Sec. 7-2. With the application of bias voltage to the substrate, these junctions represent reverse-biased diodes with a very high back resistance, thus providing adequate dc isolation. But since each *p-n* junction is also a capacitance, there remains that inevitable capacitive coupling between components and the substrate. These parasitic distributed capacitances thus limit monolithic integrated circuits to frequencies somewhat below those at which corresponding discrete circuits can operate.

Additional methods for achieving better isolation, and therefore improved frequency response, have been developed, and are discussed in this section.

Dielectric Isolation In this process[11] the diode-isolation concept is discarded completely. Instead, isolation, both electrical and physical, is achieved by means of a layer of solid dielectric which completely surrounds and separates the components from each other and from the common substrate. This passive layer can be silicon dioxide, silicon monoxide, ruby, or possibly a glazed ceramic substrate which is made thick enough so that its associated capacitance is negligible.

In a dielectric isolated integrated circuit it is possible to fabricate readily *p-n-p* and *n-p-n* transistors within the same silicon substrate. It is also simple to have both fast and charge-storage diodes and also both high- and low-frequency transistors in the same chip through selective gold diffusion—a process prohibited by conventional techniques because of the rapid rate at which gold diffuses through silicon unless impeded by a physical barrier such as a dielectric layer.

An isolation method pioneered by RCA[11] is referred to as SOS (silicon-on-sapphire). On a single-crystal sapphire substrate an *n*-type silicon layer is grown heteroepitaxially. By etching away selected portions of the silicon, isolated islands are formed (interconnected only by the high-resistance sapphire substrate).

One isolation method employing silicon dioxide as the isolating material is the EPIC process,[12] developed by Motorola, Inc. This EPIC isolation method reduces parasitic capacitance by a factor of 10 or more. In addition, the insulating oxide precludes the need for a reverse bias between substrate and circuit elements. Breakdown voltage between circuit elements and substrate is in excess of 1,000 V, in contrast to the 20 V across an isolation junction.

Beam Leads The beam-lead concept[13] of Bell Telephone Laboratories was primarily developed to batch-fabricate semiconductor devices and integrated circuits. This technique consists in depositing an array of thick (of the order of 1 mil) contacts on the surface of a slice of standard monolithic circuit, and then removing the excess semiconductor from under the contacts, thereby separating the individual devices and leaving them with semirigid beam leads cantilevered beyond the semiconductor. The contacts serve not only as electrical leads, but also as the structural support for the devices;

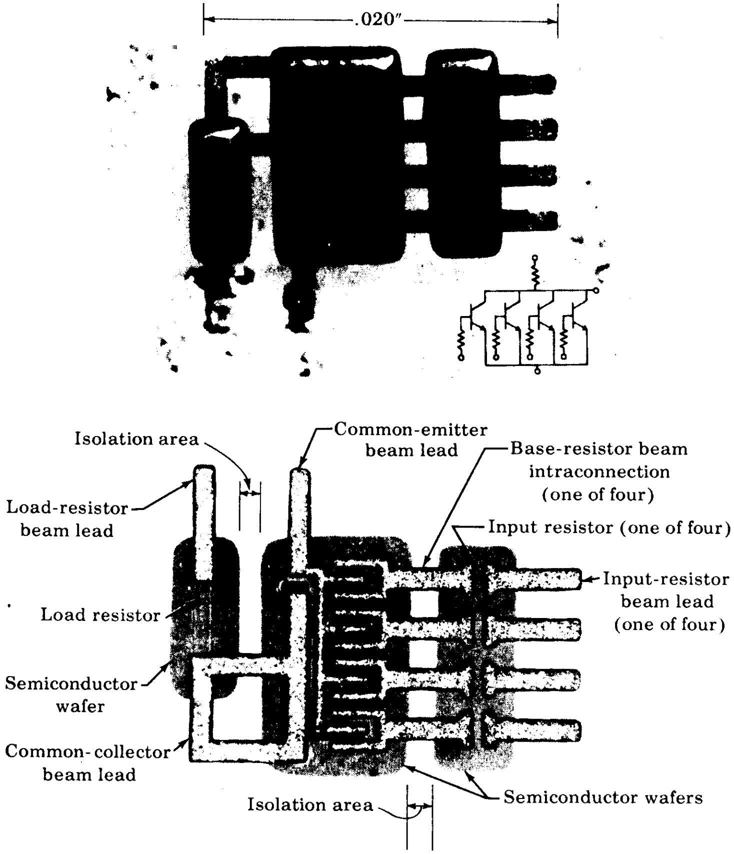

.020"
Isolation area
Common-emitter beam lead
Base-resistor beam intraconnection (one of four)
Load-resistor beam lead
Input resistor (one of four)
Input-resistor beam lead (one of four)
Load resistor
Semiconductor wafer
Common-collector beam lead
Isolation area
Semiconductor wafers

Fig. 7-28 The beam-lead isolation technique. (a) Photomicrograph of logic circuit connected in a header. (b) The underside of the same circuit, with the various elements identified. (Courtesy of Bell Telephone Laboratories.)

hence the name beam leads. Chips of beam-lead circuits are mounted directly by leads, without 1-mil aluminum or gold wires.

Isolation within integrated circuits may be accomplished by the beam-lead structure. By etching away the unwanted silicon from under the beam leads which connect the devices on an integrated chip, isolated pads of silicon may be attained, interconnected by the beam leads. The only capacitive coupling between elements is then through the small metal-over-oxide overlay. This is much lower than the junction capacitance incurred with *p-n* junction-isolated monolithic circuits.

It should be pointed out that the dielectric and beam-lead isolation techniques involve additional process steps, and thus higher costs and possible reduction in yield of the manufacturing process.

Figure 7-28 shows photomicrographs of two different views of a logic circuit made using the beam-lead technique. The top photo shows the logic circuit connected in a header. The bottom photo shows the underside of the same circuit with the various elements identified. This device is made using conventional planar techniques to form the transistor and resistor regions. Electrical isolation is accomplished by removing all unwanted material between components. The beam leads then remain to support and intraconnect the isolated components.

Hybrid Circuits[1] The hybrid circuit as opposed to the monolithic circuit consists of several component parts (transistors, diodes, resistors, capacitors, or complete monolithic circuits), all attached to the same ceramic substrate and employing wire bonding to achieve the interconnections. In these circuits electrical isolation is provided by the physical separation of the component parts, and in this respect hybrid circuits resemble beam-lead circuits.

7-12 LARGE-SCALE AND MEDIUM-SCALE INTEGRATION (LSI AND MSI)

Large-scale integration[11] represents the process of fabricating chips with a large number of components which are interconnected to form complete subsystems or systems. In 1972 commercially available LSI circuits contained, typically, more than 100 gates, or 1,000 individual circuit components. A triple-diffused bipolar transistor requires approximately 50 mil^2 of chip area, whereas a typical MOS transistor (Chap. 10) requires only 5 mil^2. Much higher element densities are possible with MOS LSI than bipolar LSI circuits.

Since LSI is an extension of integrated-circuit techniques, the fabrication is identical with that described in Sec. 7-2. Only the methods of testing and interconnection are different with LSI. There are two principal techniques, called *discretionary wiring* and *fixed interconnection pattern.* The former consists in manufacturing on a single large chip many identical units, called *unit cells,* such as logic gates which are to be interconnected into a system. The cells are then tested by an automatic LSI tester which remembers the location of the "good" cells. The tester is coupled to a digital computer which calculates instructions for a pattern of metalization runs which interconnects the good cells so as to yield the desired system function. This process must be repeated for each LSI wafer, since the patterns of good and bad circuits will differ from wafer to wafer.

A *fixed interconnection pattern* starts with a more complex cell, called a *polycell,* and then interconnects several of these to form a larger system through a fixed interconnection which is less complex than the pattern required for an equal array composed of simpler circuits.

The most common LSI products are read-write memories (R/W), read-only memories (ROM), and shift registers (discussed in Chap. 17).

Medium-scale Integration MSI devices have a component density less than that of LSI, but in excess of about 100 per chip. These commercially available units include shift registers, counters, decoders, adders, etc. (Chap. 17).

7-13 THE METAL–SEMICONDUCTOR CONTACT[14]

Two types of metal-semiconductor junctions are possible, *ohmic* and *rectifying*. The former is the type of contact desired when a lead is to be attached to a semiconductor. On the other hand, the rectifying contact results in a metal-semiconductor diode (called a *Schottky barrier*), with volt-ampere characteristics very similar to those of a p-n diode. The metal-semiconductor diode was investigated many years ago, but until the late 1960s commercial Schottky diodes were not available because of problems encountered in their manufacture. It has turned out that most of the fabrication difficulties are due to surface effects; by employing the surface-passivated integrated-circuit techniques described in this chapter, it is possible to construct almost ideal metal-semiconductor diodes very economically.

As mentioned in Sec. 7-2 (step 4), aluminum acts as a p-type impurity when in contact with silicon. If Al is to be attached as a lead to n-type Si, an ohmic contact is desired and the formation of a p-n junction must be prevented. It is for this reason that n^+ diffusions are made in the n regions near the surface where the Al is deposited (Fig. 7-2d). On the other hand, if the n^+ diffusion is omitted and the Al is deposited directly upon the n-type Si, an equivalent p-n structure is formed, resulting in an excellent metal-semiconductor diode. In Fig. 7-29 contact 1 is a Schottky barrier, whereas contact 2 is an ohmic (nonrectifying) contact, and a metal-semiconductor diode exists between these two terminals, with the anode at contact 1. Note that

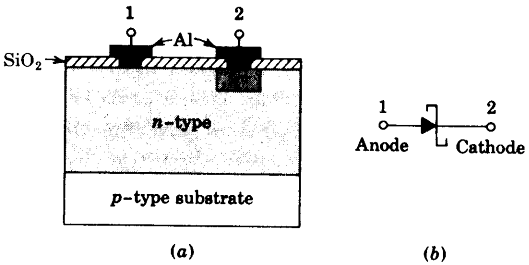


Fig. 7-29 (*a*) A Schottky diode formed by IC techniques. The aluminum and the lightly doped n region form a rectifying contact 1, whereas the metal and the heavily doped n^+ region form an ohmic contact 2. (*b*) The symbol for this metal-semiconductor diode.

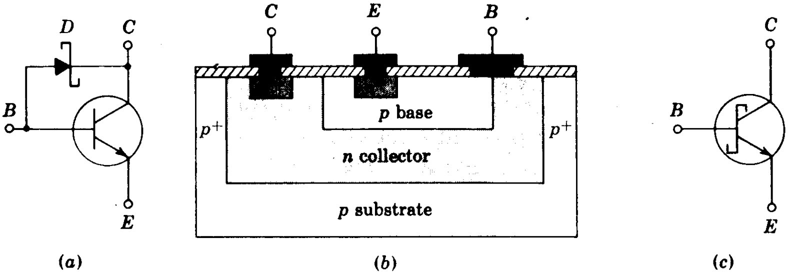


Fig. 7-30 (a) A transistor with a Schottky-diode clamp between base and collector to prevent saturation. (b) The cross section of a monolithic IC equivalent to the diode-transistor combination in (a). (c) The Schottky transistor symbol, which is an abbreviation for that shown in (a).

the fabrication of a Schottky diode is actually simpler than that of a *p-n* diode, which requires an extra (*p*-type) diffusion.

The external volt-ampere characteristic of a metal-semiconductor diode is essentially the same as that of a *p-n* junction, but the physical mechanisms involved are more complicated. Note that in the forward direction electrons from the *n*-type Si cross the junction into the metal, where electrons are plentiful. In this sense, this is a majority-carrier device, whereas minority carriers account for a *p-n* diode characteristic. As explained in Sec. 3-10, there is a delay in switching a *p-n* diode from ON to OFF because the minority carriers stored at the junction must first be removed. Schottky diodes have a negligible storage time t_s because the current is carried predominantly by majority carriers. (Electrons from the *n* side enter the aluminum and become indistinguishable from the electrons in the metal, and hence are not "stored" near the junction.)

It should be mentioned that the voltage drop across a Schottky diode is much less than that of a *p-n* diode for the same forward current. Thus, a cutin voltage of about 0.3 V is reasonable for a metal-semiconductor diode as against 0.6 V for a *p-n* barrier. Hence the former is closer to the ideal diode clamp than the latter.

The Schottky Transistor To reduce the propogation-delay time in a logic gate, it is desirable to eliminate storage time in all transistors. In other words, a transistor must be prevented from entering saturation. This condition can be achieved, as indicated in Fig. 7-30*a*, by using a Schottky diode as a clamp between the base and emitter. If an attempt is made to saturate this transistor by increasing the base current, the collector voltage drops, *D* conducts, and the base-to-collector voltage is limited to about 0.4 V. Since the collector junction is forward-biased by less than the cutin voltage (≈ 0.5 V), the transistor does *not* enter saturation (Sec. 5-8).

With no additional processing steps, the Schottky clamping diode can be fabricated at the same time that the transistor is constructed. As indicated in Fig. 7-30*b*, the aluminum metalization for the base lead is allowed to make contact also with the *n*-type collector region (but without an intervening n^+ section). This simple procedure forms a metal-semiconductor diode between base and collector. The device in Fig. 7-30*b* is equivalent to the circuit of Fig. 7-30*a*. This is referred to as a *Schottky transistor*, and is represented by the symbol in Fig. 7-30*c*.

REFERENCES

1. Motorola, Inc. (R. M. Warner, Jr., and J. N. Fordemwalt, eds.): "Integrated Circuits," McGraw-Hill Book Company, New York, 1965.
2. Phillips, A. B.: Monolithic Integrated Circuits, *IEEE Spectrum*, vol. 1, no. 6, pp. 83–101, June, 1964.
3. Jahnke, E., and F. Emde: "Tables of Functions," Dover Publications, New York, 1945.
4. Hunter, L. P.: "Handbook of Semiconductor Electronics," 2d ed., sec. 8, McGraw-Hill Book Company, New York, 1962.
5. Fuller, C. S., and J. A. Ditzenberger: Diffusion of Donor and Acceptor Elements in Silicon, *J. Appl. Phys.*, vol. 27, pp. 544–553, May, 1956.
   Barrer, P. M.: "Diffusion in and through Solids," Cambridge University Press, London, 1951.
6. Trumbore, F. A.: Solid Solubilities of Impurity Elements in Germanium and Silicon, *Bell System Tech. J.*, vol. 39, pp. 205–234, January, 1960.
7. King, D., and L. Stern: Designing Monolithic Integrated Circuits, *Semicond. Prod. Solid State Technol.*, March, 1965.
8. "Custom Microcircuit Design Handbook," Fairchild Semiconductor, Mountain View, Calif., 1963.
9. Hunter, L. P., Ref. 4, sec. 10.1.
10. Phillips, A. B.: Designing Digital Monolithic Integrated Circuits, *Motorola Monitor*, vol. 2, no. 2, pp. 18–27, 1964.
11. Khambata, A. J.: "Introduction to Large-scale Integration," John Wiley & Sons, Inc., New York, 1969.
12. Epic Process Isolates Integrated Circuit Elements with Silicon Dioxide, *Electro-technol.* (*New York*), July, 1964, p. 136.
13. Lepselter, M. P., et al.: Beam Leads and Integrated Circuits, *Proc. IEEE*, vol. 53, p. 405, April, 1965.

Lepselter, M. P.: Beam-lead Technology, *Bell System Tech. J.*, February, 1966, pp. 233–253.

14. Yu, A. Y. C.: The Metal-semiconductor Contact: An Old Device with a New Future, *IEEE Spectrum*, vol. 7, no. 3, pp. 83–89, March, 1970.

REVIEW QUESTIONS

7-1 What are the four advantages of integrated circuits?

7-2 List the five steps involved in fabricating a monolithic integrated circuit (IC), assuming you already have a substrate.

7-3 List the five basic processes involved in the fabrication of an IC, assuming you already have a substrate.

7-4 Describe *epitaxial growth.*

7-5 (*a*) Describe the *photoetching process.* (*b*) How many masks are required to complete an IC? List the function performed by each mask.

7-6 (*a*) Describe the *diffusion process.* (*b*) What is meant by an *impurity profile?*

7-7 (*a*) How is the surface layer of SiO_2 formed? (*b*) How thick is this layer? (*c*) What are the reasons for forming the SiO_2 layers?

7-8 Explain how isolation between components is obtained in an IC.

7-9 How are the components interconnected in an IC?

7-10 Explain what is meant by *parasitic capacitance* in an IC.

7-11 Give the order of magnitude of (*a*) the substrate thickness; (*b*) the epitaxial thickness; (*c*) the base width; (*d*) the diffusion time; (*e*) the diffusion temperature; (*f*) the surface area of a transistor; (*g*) the chip size.

7-12 Sketch the cross section of an IC transistor.

7-13 Sketch the cross section of a discrete planar epitaxial transistor.

7-14 List the advantages and disadvantages of an IC vs. a discrete transistor.

7-15 (*a*) Define *buried layer.* (*b*) Why is it used?

7-16 Describe a *lateral p-n-p* transistor. Why is its current gain low?

7-17 Describe a *vertical p-n-p* transistor. Why is it of limited use?

7-18 Describe a *supergain* transistor.

7-19 (*a*) How are IC diodes fabricated? (*b*) Sketch the cross sections of two types of emitter-base diodes.

7-20 Sketch the cross section of a diode pair using collector-base regions if (*a*) the cathode is common and (*b*) the anode is common.

7-21 Sketch the top view of a multiple-emitter transistor. Show the isolation, collector, base, and emitter regions.

7-22 (*a*) Define *sheet resistance* R_S. (*b*) What is the order of magnitude of R_S for the base region and also for the emitter region? (*c*) Sketch the cross section of an IC resistor. (*d*) What are the order of magnitudes of the smallest and the largest values of an IC resistance?

7-23 (*a*) Sketch the equivalent circuit of a base-diffused resistor, showing all parasitic elements. (*b*) What must be done (externally) to minimize the effect of the parasitic elements?

7-24 Describe a *thin-film resistor.*

7-25 (*a*) Sketch the cross section of a junction capacitor. (*b*) Draw the equivalent circuit, showing all parasitic elements.

7-26 Repeat Rev. 7-25 for an MOS capacitor.

7-27 (*a*) What are the two basic distinctions between a junction and an MOS capacitor? (*b*) What is the order of magnitude of the capacitance per square mil? (*c*) What is the order of magnitude of the maximum value of C?

7-28 (*a*) To what voltage is the substrate connected? Why? (*b*) To what voltage is the isolation island containing the resistors connected? Why? (*c*) Can several transistors be placed in the same isolation island? Explain.

7-29 Describe briefly (*a*) *dielectric isolation;* (*b*) *beam-lead isolation.*

7-30 What is meant by a *hybrid circuit?*

7-31 How is an aluminum contact made with *n*-type silicon so that it is (*a*) ohmic; (*b*) rectifying?

7-32 Why is storage time eliminated in a metal-semiconductor diode?

7-33 What is a *Schottky transistor?* Why is storage time eliminated in such a transistor?

7-34 Sketch the cross section of an IC Schottky transistor.

10/ FIELD-EFFECT TRANSISTORS

The field-effect transistor[1] is a semiconductor device which depends for its operation on the control of current by an electric field. There are two types of field-effect transistors, the *junction field-effect transistor* (abbreviated JFET, or simply FET) and the *insulated-gate field-effect transistor* (IGFET), more commonly called the *metal-oxide-semiconductor (MOS) transistor* (MOST, or MOSFET).

The principles on which these devices operate, as well as the differences in their characteristics, are examined in this chapter. Representative circuits making use of FET transistors are also presented.

The field-effect transistor differs from the bipolar junction transistor in the following important characteristics:

1. Its operation depends upon the flow of majority carriers only. It is therefore a *unipolar* (one type of carrier) device.
2. It is simpler to fabricate and occupies less space in integrated form.
3. It exhibits a high input resistance, typically many megohms.
4. It is less noisy than a bipolar transistor.
5. It exhibits no offset voltage at zero drain current, and hence makes an excellent signal chopper.[2]

The main disadvantage of the FET is its relatively small gain-bandwidth product in comparison with that which can be obtained with a conventional transistor. The principal applications of MOSFETs are as LSI digital arrays.

10-1 THE JUNCTION FIELD–EFFECT TRANSISTOR

The structure of an *n-channel* field-effect transistor is shown in Fig. 10-1. Ohmic contacts are made to the two ends of a semiconductor bar of *n*-type material (if *p*-type silicon is used, the device is referred to as a *p-channel* FET). Current is caused to flow along the length of the bar because of the voltage supply connected between the ends. This current consists of majority carriers, which in this case are electrons. A simple side view of a JFET is indicated in Fig. 10-1*a* and a more detailed sketch is shown in Fig. 10-1*b*. The circuit symbol with current and voltage polarities marked is given in Fig. 10-2. The following FET notation is standard.

Source The *source* S is the terminal through which the majority carriers enter the bar. Conventional current entering the bar at S is designated by I_S.

Drain The *drain* D is the terminal through which the majority carriers leave the bar. Conventional current entering the bar at D is designated by I_D. The drain-to-source voltage is called V_{DS}, and is positive if D is more positive than S. In Fig. 10-1, $V_{DS} = V_{DD}$ = drain supply voltage.

Gate On both sides of the n-type bar of Fig. 10-1, heavily doped (p^+) regions of acceptor impurities have been formed by alloying, by diffusion, or

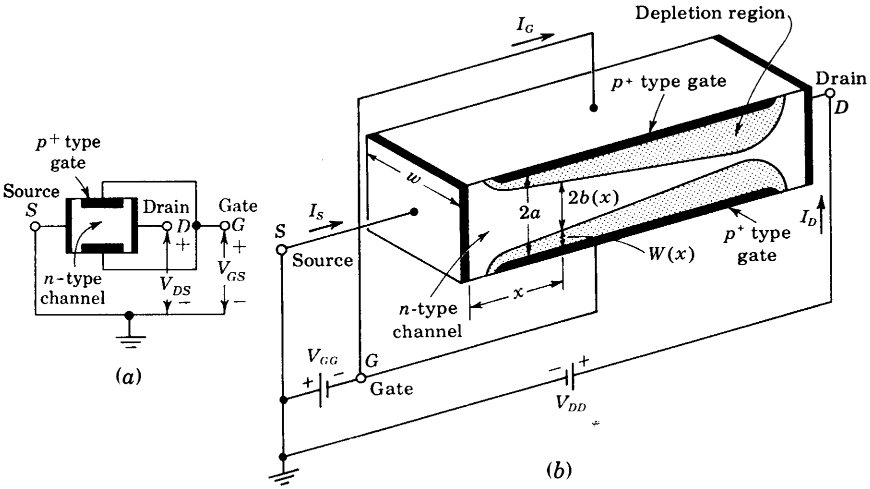


Fig. 10-1 The basic structure of an n-channel field-effect transistor. (a) Simplified view. (b) More detailed drawing. The normal polarities of the drain-to-source (V_{DD}) and gate-to-source (V_{GG}) supply voltages are shown. In a p-channel FET the voltages would be reversed.

by any other procedure available for creating *p-n* junctions. These impurity regions are called the *gate G*. Between the gate and source a voltage $V_{GS} = -V_{GG}$ is applied in the direction to reverse-bias the *p-n* junction. Conventional current entering the bar at G is designated I_G.

Channel The region in Fig. 10-1 of n-type material between the two gate regions is the *channel* through which the majority carriers move from source to drain.

FET Operation It is necessary to recall that on the two sides of the reverse-biased *p-n* junction (the transition region) there are space-charge regions (Sec. 3-7). The current carriers have diffused across the junction, leaving only uncovered positive ions on the n side and negative ions on the p side. The electric lines of field intensity which now originate on the positive ions and terminate on the negative ions are precisely the source of the voltage drop across the junction. As the reverse bias across the junction increases, so also does the thickness of the region of immobile uncovered charges. The conductivity of this region is nominally zero because of the unavailability of current carriers. Hence we see that the effective width of the *channel* in Fig. 10-1 will become progressively decreased with increasing reverse bias. Accordingly, for a fixed drain-to-source voltage, the drain current will be a function of the reverse-biasing voltage across the gate junction. The term *field effect* is used to describe this device because the mechanism of current control is the *effect* of the extension, with increasing reverse bias, of the *field* associated with the region of uncovered charges.

FET Static Characteristics The circuit, symbol, and polarity conventions for an FET are indicated in Fig. 10-2. The direction of the arrow at the gate of the junction FET in Fig. 10-2 indicates the direction in which gate current would flow if the gate junction were forward-biased. The common-source drain characteristics for a typical n-channel FET shown in Fig. 10-3 give I_D against V_{DS}, with V_{GS} as a parameter. To see qualitatively why the characteristics have the form shown, consider, say, the case for which $V_{GS} = 0$. For $I_D = 0$, the channel between the gate junctions is entirely open. In response

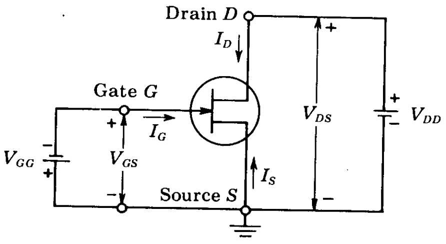


Fig. 10-2 Circuit symbol for an n-channel FET. (For a p-channel FET the arrow at the gate junction points in the opposite direction.) For an n-channel FET, I_D and V_{DS} are positive and V_{GS} is negative. For a p-channel FET, I_D and V_{DS} are negative and V_{GS} is positive.

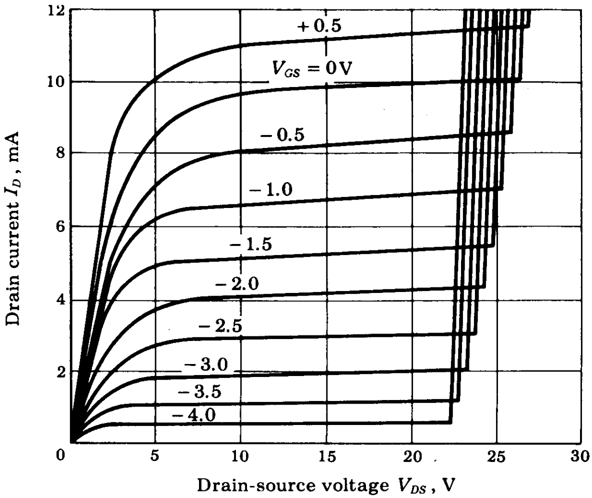


Fig. 10-3 Common-source drain characteristics of an n-channel field-effect transistor.

to a small applied voltage V_{DS}, the n-type bar acts as a simple semiconductor resistor, and the current I_D increases linearly with V_{DS}. With increasing current, the ohmic voltage drop between the source and the channel region reverse-biases the junction, and the conducting portion of the channel begins to constrict. Because of the ohmic drop along the length of the channel itself, the constriction is not uniform, but is more pronounced at distances farther from the source, as indicated in Fig. 10-1. Eventually, a voltage V_{DS} is reached at which the channel is "pinched off." This is the voltage, not too sharply defined in Fig. 10-3, where the current I_D begins to level off and approach a constant value. It is, of course, in principle not possible for the channel to close completely and thereby reduce the current I_D to zero. For if such, indeed, could be the case, the ohmic drop required to provide the necessary back bias would itself be lacking. Note that each characteristic curve has an ohmic region for small values of V_{DS}, where I_D is proportional to V_{DS}. Each also has a constant-current region for large values of V_{DS}, where I_D responds very slightly to V_{DS}.

If now a gate voltage V_{GS} is applied in the direction to provide additional reverse bias, pinch-off will occur for smaller values of $|V_{DS}|$, and the maximum drain current will be smaller. This feature is brought out in Fig. 10-3. Note that a plot for a silicon FET is given even for $V_{GS} = +0.5$ V, which is in the direction of forward bias. We note from Table 5-1 that, actually, the gate current will be very small, because at this gate voltage the Si junction is barely at the cutin voltage V_γ.

The maximum voltage that can be applied between any two terminals of the FET is the lowest voltage that will cause avalanche breakdown (Sec. 3-11) across the gate junction. From Fig. 10-3 it is seen that avalanche occurs at a lower value of $|V_{DS}|$ when the gate is reverse-biased than for $V_{GS} = 0$. This

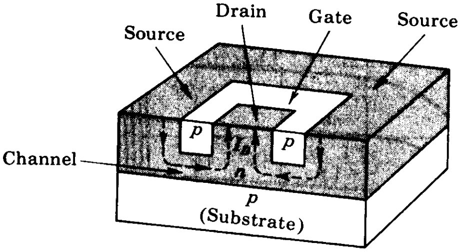


Fig. 10-4 Single-ended-geometry junction FET.

is caused by the fact that the reverse-bias gate voltage adds to the drain voltage, and hence increases the effective voltage across the gate junction.

We note from Fig. 10-2 that the n-channel FET requires zero or negative gate bias and positive drain voltage. The p-channel FET requires opposite voltage polarities. Either end of the channel may be used as a source. We can remember supply polarities by using the channel type, p or n, to designate the polarity of the *source* side of the drain supply.

A Practical FET Structure The structure shown in Fig. 10-1 is not practical because of the difficulties involved in diffusing impurities into both sides of a semiconductor wafer. Figure 10-4 shows a single-ended-geometry junction FET where diffusion is from one side only. The substrate is of p-type material onto which an n-type channel is epitaxially grown (Sec. 7-3). A p-type gate is then diffused into the n-type channel. The substrate which may function as a second gate is of relatively low resistivity material. The diffused gate is also of very low resistivity material, allowing the depletion region to spread mostly into the n-type channel.

10-2 THE PINCH-OFF VOLTAGE V_P

We derive an expression for the gate reverse voltage V_P that removes all the free charge from the channel using the physical model described in the preceding section. This analysis was first made by Shockley,[1] using the structure of Fig. 10-1. In this device a slab of n-type semiconductor is sandwiched between two layers of p-type material, forming two p-n junctions.

Assume that the p-type region is doped with N_A acceptors per cubic meter, that the n-type region is doped with N_D donors per cubic meter, and that the junction formed is abrupt. The assumption of an abrupt junction is the same as that made in Sec. 3-7 and Fig. 3-10, and is chosen for simplicity. Moreover, if $N_A \gg N_D$, we see from Eq. (3-17) that $W_p \ll W_n$, and using Eq. (3-21), we have, for the space-charge width, $W_n(x) = W(x)$ at a distance x along the channel in Fig. 10-1:

$$W(x) = a - b(x) = \left\{\frac{2\epsilon}{qN_D}[V_o - V(x)]\right\}^{\frac{1}{2}} \tag{10-1}$$

where ϵ = dielectric constant of channel material
q = magnitude of electronic charge
V_o = junction contact potential at x (Fig. 3-1d)
$V(x)$ = applied potential across space-charge region at x and is a negative number for an applied reverse bias
$a - b(x)$ = penetration $W(x)$ of depletion region into channel at a point x along channel (Fig. 10-1)

If the drain current is zero, $b(x)$ and $V(x)$ are independent of x and $b(x) = b$. If in Eq. (10-1) we substitute $b(x) = b = 0$ and solve for V, on the assumption that $|V_o| \ll |V|$, we obtain the pinch-off voltage V_P, the diode reverse voltage that removes all the free charge from the channel. Hence

$$|V_P| = \frac{qN_D}{2\epsilon} a^2 \tag{10-2}$$

If we substitute V_{GS} for $V_o - V(x)$ in Eq. (10-1), we obtain, using Eq. (10-2),

$$V_{GS} = \left(1 - \frac{b}{a}\right)^2 V_P \tag{10-3}$$

The voltage V_{GS} in Eq. (10-3) represents the reverse bias across the gate junction and is independent of distance along the channel if $I_D = 0$.

EXAMPLE For an n-channel silicon FET with $a = 3 \times 10^{-4}$ cm and $N_D = 10^{15}$ electrons/cm^3, find (a) the pinch-off voltage and (b) the channel half-width for $V_{GS} = \frac{1}{2}V_P$ and $I_D = 0$.

Solution a. The relative dielectric constant of silicon is given in Table 2-1 as 12, and hence $\epsilon = 12\epsilon_o$. Using the values of q and ϵ_o from Appendix A, we have, from Eq. (10-2), expressed in mks units,

$$V_P = \frac{1.60 \times 10^{-19} \times 10^{21} \times (3 \times 10^{-6})^2}{2 \times 12 \times (36\pi \times 10^9)^{-1}} = 6.8 \text{ V}$$

b. Solving Eq. (10-3) for b, we obtain for $V_{GS} = \frac{1}{2}V_P$

$$b = a\left[1 - \left(\frac{V_{GS}}{V_P}\right)^{\frac{1}{2}}\right] = (3 \times 10^{-4})[1 - (\tfrac{1}{2})^{\frac{1}{2}}| = 0.87 \times 10^{-4} \text{ cm}$$

Hence the channel width has been reduced to about one-third its value for $V_{GS} = 0$.

10-3 THE JFET VOLT–AMPERE CHARACTERISTICS

Assume, first, that a small voltage V_{DS} is applied between drain and source. The resulting small drain current I_D will then have no appreciable effect on the channel profile. Under these conditions we may consider the effective channel cross section A to be constant throughout its length. Hence $A = 2bw$,

orresponding to zero drain current as given by
and w is the channel dimension perpendicular
n Fig. 10-1.
the depletion region, then, using Ohm's law
ain current

$$N_D\mu_n \frac{V_{DS}}{L} \qquad (10\text{-}4)$$

nnel.
-3) in Eq. (10-4), we have

$$\left.\frac{V_{GS}}{V_P}\right)^{1/2}\Bigg] V_{DS} \qquad (10\text{-}5)$$

Equation (10-5) describes the volt-ampere
ery small V_{DS}, and it suggests that under these
ke an ohmic resistance whose value is deter-
$/I_D$ at the origin is called the ON *drain resis-*
tain from Eq. (10-5), with $V_{GS} = 0$,

(10-6)

the illustrative example in this section and
$_N = 3.3$ K. For the dimensions and concen-
vailable FETs and MOSFETs (Sec. 10-5),
t 100 Ω to 100 K are measured. This param-
plications where the FET is driven heavily ON.
dvantage over the field-effect device in that
and hence is much smaller than $r_{d,\text{ON}}$. How-
e disadvantage for chopper applications[2] of
. 5-12), whereas the FET characteristics pass
$V_{DS} = 0$.

now consider the situation where an electric
If a substantial drain current I_D flows, the
verse-biased than the source end, and hence
egion are not parallel to the longitudinal axis
hown in Fig. 10-1. If the convergence of the
revious one-dimensional analysis is valid[1] in a
ness Δx and at a distance x from the source.
gradual" channel, the current may be written

(10-7)

increase, whereas $b(x)$ decreases because the
rrent density $J = I_D/2b(x)w$ increases. We

now see that complete pinch-off ($b = 0$) cannot take place because, if it did, J would become infinite, which is a physically impossible condition. If J were to increase without limit, then, from Eq. (10-7), so also would $\mathcal{E}_x$, provided that μ_n remains constant. It is found experimentally,[3,4] however, that the mobility is a function of electric field intensity and remains constant only for $\mathcal{E}_x < 10^3$ V/cm in n-type silicon. For moderate fields, 10^3 to 10^4 V/cm, the mobility is approximately inversely proportional to the square root of the applied field. For still higher fields, such as are encountered at pinch-off, μ_n is inversely proportional to $\mathcal{E}_x$. In this region the drift velocity of the electrons ($v_x = \mu_n \mathcal{E}_x$) remains constant, and Ohm's law is no longer valid. From Eq. (10-7) we now see that both I_D and b remain constant, thus explaining the constant-current portion of the V-I characteristic of Fig. 10-3.

What happens[4] if V_{DS} is increased beyond pinch-off, with V_{GS} held constant? As explained above, the minimum channel width $b_{\min} = \delta$ has a small nonzero constant value. This minimum width occurs at the drain end of the bar. As V_{DS} is increased, this increment in potential causes an increase in $\mathcal{E}_x$ in an adjacent channel section toward the source. Referring to Fig. 10-5, the velocity-limited region L' increases with V_{DS}, whereas δ remains at a fixed value.

The Region before Pinch-off We have verified that the FET behaves as an ohmic resistance for small V_{DS} and as a constant-current device for large V_{DS}. An analysis giving the shape of the volt-ampere characteristic between these two extremes is complicated. It has already been mentioned that in this region the mobility is at first independent of electric field and then μ varies with $\mathcal{E}_x^{-\frac{1}{2}}$ for larger values of $\mathcal{E}_x$ (before pinch-off). Taking this relationship into account, it is possible[3-5] to obtain an expression for I_D as a

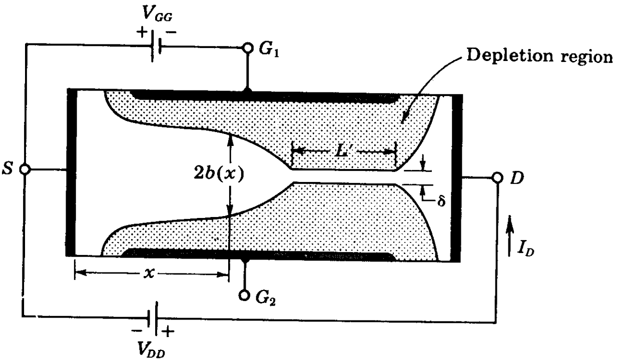


Fig. 10-5 After pinch-off, as V_{DS} is increased, then L' increases but δ and I_D remain essentially constant. (G_1 and G_2 are tied together.)

function of V_{DS} and V_{GS} which agrees quite well with experimentally determined curves.

The Transfer Characteristic In amplifier applications the FET is almost always used in the region beyond pinch-off (also called the *constant-current, pentode,* or *current-saturation region*). Let the saturation drain current be designated by I_{DS}, and its value with the gate shorted to the source ($V_{GS} = 0$) by I_{DSS}. It has been found[6] that the transfer characteristic, giving the relationship between I_{DS} and V_{GS}, can be approximated by the parabola

$$I_{DS} = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{10-8}$$

This simple parabolic approximation gives an excellent fit, with the experimentally determined transfer characteristics for FETs made by the diffusion process.

Cutoff Consider an FET operating at a fixed value of V_{DS} in the constant-current region. As V_{GS} is increased in the direction to reverse-bias the gate junction, the conducting channel will narrow. When $V_{GS} = V_P$, the channel width is reduced to zero, and from Eq. (10-8), $I_{DS} = 0$. With a physical device some leakage current $I_{D,\mathrm{OFF}}$ still flows even under the cutoff condition $|V_{GS}| > |V_P|$. A manufacturer usually specifies a maximum value of $I_{D,\mathrm{OFF}}$ at a given value of V_{GS} and V_{DS}. Typically, a value of a few nanoamperes may be expected for $I_{D,\mathrm{OFF}}$ for a silicon FET.

The gate reverse current, also called *the gate cutoff current,* designated by I_{GSS}, gives the gate-to-source current, with the drain shorted to the source for $|V_{GS}| > |V_P|$. Typically, I_{GSS} is of the order of a few nanoamperes for a silicon device.

10-4 THE FET SMALL-SIGNAL MODEL

The linear small-signal equivalent circuit for the FET can be obtained in a manner analogous to that used to derive the corresponding model for a transistor. We employ the same notation in labeling time-varying and dc currents and voltages as used in Secs. 8-1 and 8-2 for the transistor. We can formally express the drain current i_D as a function f of the gate voltage v_{GS} and drain voltage v_{DS} by

$$i_D = f(v_{GS}, v_{DS}) \tag{10-9}$$

The Transconductance g_m and Drain Resistance r_d If both the gate and drain voltages are varied, the change in drain current is given approximately by the first two terms in the Taylor's series expansion of Eq. (10-9), or

$$\Delta i_D = \left.\frac{\partial i_D}{\partial v_{GS}}\right|_{V_{DS}} \Delta v_{GS} + \left.\frac{\partial i_D}{\partial v_{DS}}\right|_{V_{GS}} \Delta v_{DS} \tag{10-10}$$

In the small-signal notation of Sec. 8-1, $\Delta i_D = i_d$, $\Delta v_{GS} = v_{gs}$, and $\Delta v_{DS} = v_{ds}$, so that Eq. (10-10) becomes

$$i_d = g_m v_{gs} + \frac{1}{r_d} v_{ds} \tag{10-11}$$

where

$$g_m \equiv \left.\frac{\partial i_D}{\partial v_{GS}}\right|_{V_{DS}} \approx \left.\frac{\Delta i_D}{\Delta v_{GS}}\right|_{V_{DS}} = \left.\frac{i_d}{v_{gs}}\right|_{V_{DS}} \tag{10-12}$$

is the *mutual conductance*, or *transconductance.* It is also often designated by y_{fs} or g_{fs} and called the (*common-source*) *forward transadmittance.* The second parameter r_d in Eq. (10-11) is the *drain* (or *output*) *resistance,* and is defined by

$$r_d \equiv \left.\frac{\partial v_{DS}}{\partial i_D}\right|_{V_{GS}} \approx \left.\frac{\Delta v_{DS}}{\Delta i_D}\right|_{V_{GS}} = \left.\frac{v_{ds}}{i_d}\right|_{V_{GS}} \tag{10-13}$$

The reciprocal of r_d is the drain conductance g_d. It is also designated by y_{os} and g_{os} and called the (common-source) output conductance.

An *amplification factor* μ for an FET may be defined by

$$\mu \equiv -\left.\frac{\partial v_{DS}}{\partial v_{GS}}\right|_{I_D} = -\left.\frac{\Delta v_{DS}}{\Delta v_{GS}}\right|_{I_D} = -\left.\frac{v_{ds}}{v_{gs}}\right|_{i_d=0} \tag{10-14}$$

We can verify that μ, r_d, and g_m are related by

$$\mu = r_d g_m \tag{10-15}$$

by setting $i_d = 0$ in Eq. (10-11).

An expression for g_m is obtained by applying the definition of Eq. (10-12) to Eq. (10-8). The result is

$$g_m = g_{mo}\left(1 - \frac{V_{GS}}{V_P}\right) = -\frac{2}{V_P}(I_{DSS} I_{DS})^{\frac{1}{2}} \tag{10-16}$$

where g_{mo} is the value of g_m for $V_{GS} = 0$, and is given by

$$g_{mo} = \frac{-2I_{DSS}}{V_P} \tag{10-17}$$

Since I_{DSS} and V_P are of opposite sign, g_{mo} is always positive. Note that the transconductance varies as the square root of the drain current. The relationship connecting g_{mo}, I_{DSS}, and V_P has been verified experimentally.[7] Since g_{mo} can be measured and I_{DSS} can be read on a dc milliammeter placed in the drain lead (with zero gate excitation), Eq. (10-17) gives a method for obtaining V_P.

The dependence of g_m upon V_{GS} is indicated in Fig. 10-6 for the 2N3277 FET (with $V_P \approx 4.5$ V) and the 2N3278 FET (with $V_P \approx 7$ V). The linear relationship predicted by Eq. (10-16) is seen to be only approximately valid.

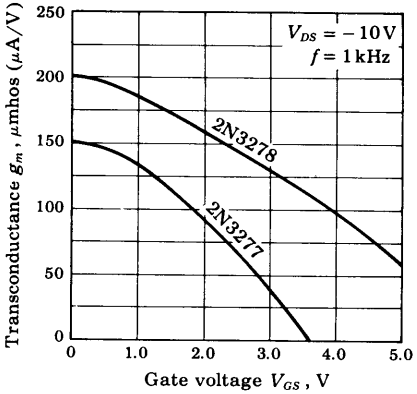


Fig. 10-6 Transconductance g_m versus gate voltage for types 2N3277 and 2N3278 FETs. (Courtesy of Fairchild Semiconductor Company.)

Temperature Dependence Curves of g_m and r_d versus temperature are given in Fig. 10-7. The drain current I_{DS} has the same temperature variation as does g_m. The principal reason for the negative temperature coefficient of I_{DS} is that the mobility decreases with increasing temperature.[8] Since this majority-carrier current decreases with temperature (unlike the bipolar transistor whose minority-carrier current increases with temperature), the troublesome phenomenon of *thermal runaway* (Sec. 9-9) is not encountered with field-effect transistors.

The FET Model A circuit which satisfies Eq. (10-11) is indicated in Fig. 10-8*a*. This low-frequency small-signal model has a Norton's output circuit with a dependent current generator whose current is proportional to the gate-to-source voltage. The proportionality factor is the transconductance g_m, which is consistent with the definition of g_m in Eq. (10-12). The output resistance is r_d, which is consistent with the definition in Eq. (10-13). The input resistance between gate and source is infinite, since it is assumed that

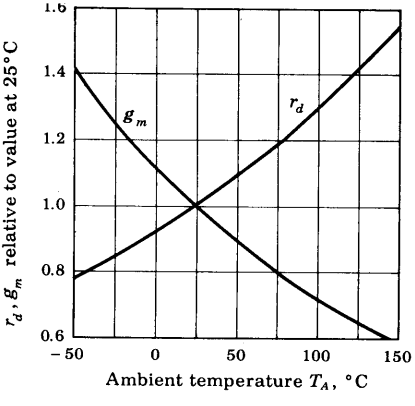


Fig. 10-7 Normalized g_m and normalized r_d versus T_A (for the 2N3277 and the 2N3278 FETs with $V_{DS} = -10$ V, $V_{GS} = 0$ V, and $f = 1$ kHz). (Courtesy of Fairchild Semiconductor Company.)

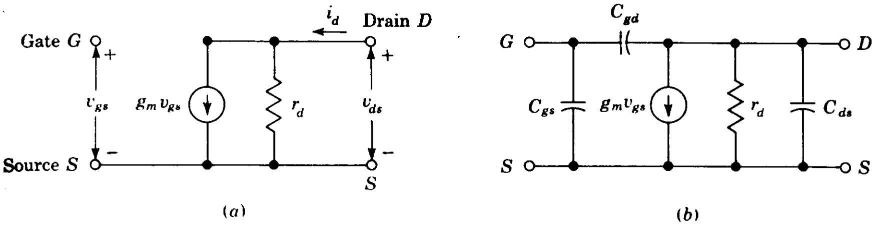


Fig. 10-8 (*a*) **The low-frequency small-signal FET model.** (*b*) **The high-frequency model, taking node capacitors into account.**

the reverse-biased gate takes no current. For the same reason the resistance between gate and drain is assumed to be infinite.

The FET model of Fig. 10-8*a* should be compared with the *h*-parameter model of the bipolar junction transistor of Fig. 8-6. The latter also has a Norton's output circuit, but the current generated depends upon the input *current*, whereas in the FET model the generator current depends upon the input *voltage*. Note that there is no feedback at low frequencies from output to input in the FET, whereas such feedback exists in the bipolar transistor through the parameter h_{re}. Finally, observe that the high (almost infinite) input resistance of the FET is replaced by an input resistance of about 1 K for a CE amplifier. In summary, the field-effect transistor is a much more ideal amplifier than the conventional transistor *at low frequencies*. Unfortunately, this is not true beyond the audio range, as we now indicate.

The high-frequency model given in Fig. 10-8*b* is identical with Fig. 10-8*a* except that the capacitances between pairs of nodes have been added. The capacitor C_{gs} represents the barrier capacitance between gate and source, and C_{gd} is the barrier capacitance between gate and drain. The element C_{ds} represents the drain-to-source capacitance of the channel. Because of these internal capacitances, feedback exists between the input and output circuits, and the voltage amplification drops rapidly as the frequency is increased (Sec. 10-11). The order of magnitudes of the parameters in the model for a diffused-junction FET is given in Table 10-1.

TABLE 10-1 **Range of parameter values for an FET**

Parameter	JFET	MOSFET†
g_m	0.1–10 mA/V	0.1–20 mA/V or more
r_d	0.1–1 M	1–50 K
C_{ds}	0.1–1 pF	0.1–1 pF
C_{gs}, C_{gd}	1–10 pF	1–10 pF
r_{gs}	$>10^8\ \Omega$	$>10^{10}\ \Omega$
r_{gd}	$>10^8\ \Omega$	$>10^{14}\ \Omega$

† Discussed in Sec. 10-5.

10-5 THE METAL-OXIDE-SEMICONDUCTOR FET (MOSFET)

In preceding sections we developed the volt-ampere characteristics and small-signal properties of the junction field-effect transistor. We now turn our attention to the insulated-gate FET, or metal-oxide-semiconductor FET,[9] which is of much greater commercial importance than the junction FET.

The p-channel MOSFET consists of a lightly doped n-type substrate into which two highly doped p^+ regions are diffused, as shown in Fig. 10-9. These p^+ sections, which will act as the source and drain, are separated by about 10 to 20 μm. A thin (1,000 to 2,000 Å) layer of insulating silicon dioxide (SiO_2) is grown over the surface of the structure, and holes are cut into the oxide layer, allowing contact with the source and drain. Then the gate-metal area is overlaid on the oxide, covering the entire channel region. Simultaneously, metal contacts are made to the drain and source, as shown in Fig. 10-9. The contact to the metal over the channel area is the gate terminal. The chip area of a MOSFET is 5 square mils or less, which is only about 5 percent of that required by a bipolar junction transistor.

The metal area of the gate, in conjunction with the insulating dielectric oxide layer and the semiconductor channel, form a parallel-plate capacitor. The insulating layer of silicon dioxide is the reason why this device is called the insulated-gate field-effect transistor. This layer results in an extremely high input resistance (10^{10} to 10^{15} Ω) for the MOSFET. The p-channel enhancement MOSFET is the most commonly available field-effect device (1972), and its characteristics will now be described.

The Enhancement MOSFET If we ground the substrate for the structure of Fig. 10-9 and apply a negative voltage at the gate, an electric field will be directed perpendicularly through the oxide. This field will end on "induced" positive charges on the semiconductor site, as shown in Fig. 10-9. The positive charges, which are minority carriers in the n-type substrate, form an "inversion layer." As the magnitude of the negative voltage on the gate increases, the induced positive charge in the semiconductor increases. The region beneath the oxide now has p-type carriers, the conductivity increases, and current flows from source to drain through the induced channel. Thus

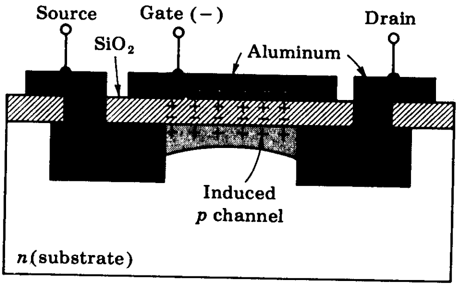


Fig. 10-9 Enhancement in a p-channel MOSFET. (Courtesy of Motorola Semiconductor Products, Inc.)

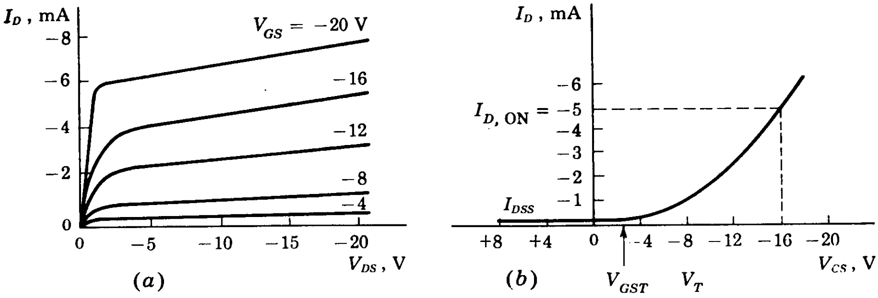


Fig. 10-10 (a) The drain characteristics and (b) the transfer curve (for V_{DS} = 10 V) of a *p*-channel enhancement-type MOSFET.

the drain current is "enhanced" by the negative gate voltage, and such a device is called an *enhancement-type* MOS.

Threshold Voltage The volt-ampere drain characteristics of a *p*-channel enhancement-mode MOSFET are given in Fig. 10-10*a*, and its transfer curve in Fig. 10-10*b*. The current I_{DSS} at $V_{GS} \geq 0$ is very small, of the order of a few nanoamperes. As V_{GS} is made negative, the current $|I_D|$ increases slowly at first, and then much more rapidly with an increase in $|V_{GS}|$. The manufacturer often indicates the *gate-source* threshold voltage V_{GST}, or V_T,† at which $|I_D|$ reaches some defined small value, say 10 μA. A current $I_{D,\mathrm{ON}}$ corresponding approximately to the maximum value given on the drain characteristics, and the value of V_{GS} needed to obtain this current are also usually given on the manufacturer's specification sheets.

The value of V_T for the *p*-channel standard MOSFET is typically −4 V, and it is common to use a power-supply voltage of −12 V for the drain supply. This large voltage is incompatible with the power-supply voltage of typically 5 V used in bipolar integrated circuits. Thus various manufacturing techniques[10] have been developed to reduce V_T. In general, a low threshold voltage allows (1) the use of a small power-supply voltage, (2) compatible operation with bipolar devices, and (3) smaller switching time due to the smaller voltage swing during switching.

Three methods are used to lower the magnitude of V_T.

1. The high-threshold MOSFET described above uses a silicon crystal with ⟨111⟩ orientation. If a crystal is utilized in the ⟨100⟩ direction it is found that a value of V_T results which is about one-half that obtained with ⟨111⟩ orientation.

2. The silicon nitride approach makes use of a layer of Si_3N_4 and SiO_2,

---

† In this chapter the threshold voltage should not be confused with the volt equivalent of temperature V_T of Sec. 2-9.

whose dielectric constant is about twice that of SiO_2 alone. A FET constructed in this manner (designated an MNOS device) decreases V_T to approximately 2 V.

3. Polycrystalline silicon doped with boron is used as the gate electrode instead of aluminum. This reduction in the difference in contact potentials between the gate electrode and the gate dielectric reduces V_T. Such devices are called *silicon gate* MOS transistors. All three of the fabrication methods described above result in a low-threshold device with V_T in the range 1.5 to 2.5 V, whereas the standard high-threshold MOS has a V_T of approximately 4 to 6 V.

Power Supply Requirements Table 10-2 gives the voltages customarily used with high-threshold and low-threshold *p*-channel MOSFETs. Note that V_{SS} refers to the substrate, V_{DD} to the drain, and V_{GG} to the gate supply voltages. The subscript 1 denotes that the source is grounded and the subscript 2 designates that the drain is at ground potential.

The low-threshold MOS circuits require lower power supply voltages and this means less expensive system power supplies. In addition, the input voltage swing for turning the device ON and OFF is smaller for the lower-threshold voltage, and this means faster operation. Another very desirable feature of low-threshold MOS circuits is that they are directly compatible with bipolar *IC*s. They require and produce essentially the same input and output signal swings and the system designer has the flexibility of using MOS and bipolar circuits in the same system.

TABLE 10-2 Power supply voltages for *p*-channel MOSFETs, in volts

	V_{SS1}	V_{DD1}	V_{GG1}	V_{SS2}	V_{DD2}	V_{GG2}
High-threshold	0	−12	−24	+12	0	−12
Low-threshold	0	−5	−17	+5	0	−12

Ion Implantation[10] The ion-implantation technique demonstrated in Fig. 10-11 provides very precise control of doping. Ions of the proper dopant such as phosphorus or boron are accelerated to a high energy of up to 300,000 eV and are used to bombard the silicon wafer target. The energy of the ions determines the depth of penetration into the target. In those areas where ion implantation is not desired, an aluminum mask or a thick (12,000 Å) oxide layer absorbs the ion. Virtually any value of V_T can be obtained using ion implantation. In addition, we see from Fig. 10-11 that there is no overlap between the gate and drain or gate and source electrodes (compare Fig. 10-11 with Fig. 10-9). Consequently, due to ion implantation, there is a drastic reduction in C_{gd} and C_{gs}.

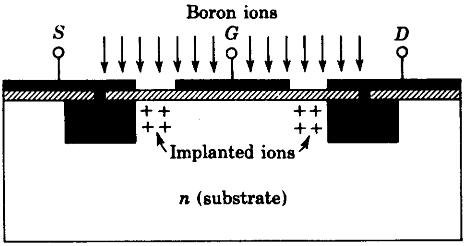


Fig. 10-11 Ion implantation in MOS devices.

The Depletion MOSFET A second type of MOSFET can be made if, to the basic structure of Fig. 10-9, a channel is diffused between the source and the drain, with the same type of impurity as used for the source and drain diffusion. Let us now consider such an n-channel structure, shown in Fig. 10-12*a*. With this device an appreciable drain current I_{DSS} flows for zero gate-to-source voltage $V_{GS} = 0$. If the gate voltage is made negative, positive charges are induced in the channel through the SiO_2 of the gate capacitor. Since the current in an FET is due to majority carriers (electrons for an n-type material), the induced positive charges make the channel less conductive, and the drain current drops as V_{GS} is made more negative. The redistribution of charge in the channel causes an effective depletion of majority carriers, which accounts for the designation *depletion* MOSFET. Note in Fig. 10-12*b* that, because of the voltage drop due to the drain current, the channel region nearest the drain is more depleted than is the volume near the source. This phenomenon is analogous to that of pinch-off occurring in a JFET at the drain end of the channel (Fig. 10-1). As a matter of fact, the volt-ampere characteristics of the depletion-mode MOS and the JFET are quite similar.

A MOSFET of the depletion type just described may also be operated

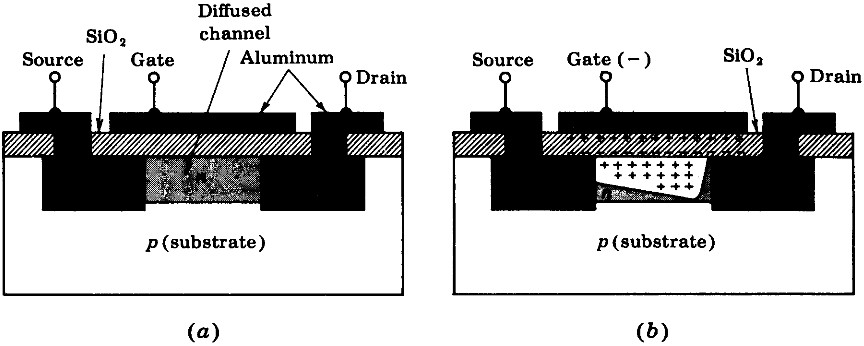


Fig. 10-12 (*a*) An n-channel depletion-type MOSFET. (*b*) Channel depletion with the application of a negative gate voltage. (Courtesy of Motorola Semiconductor Products, Inc.)

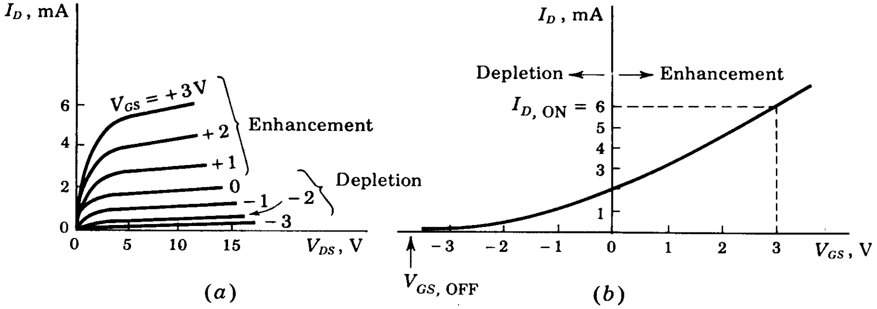


Fig. 10-13 (a) The drain characteristics and (b) the transfer curve (for $V_{DS} = 10$ V) for an n-channel MOSFET which may be used in either the enhancement or the depletion mode.

in an enhancement mode. It is only necessary to apply a positive gate voltage so that negative charges are induced into the n-type channel. In this manner the conductivity of the channel increases and the current rises above I_{DSS}. The volt-ampere characteristics of this device are indicated in Fig. 10-13a, and the transfer curve is given in Fig. 10-13b. The depletion and enhancement regions, corresponding to V_{GS} negative and positive, respectively, should be noted. The manufacturer sometimes indicates the *gate-source cutoff voltage* $V_{GS,\text{OFF}}$, at which I_D is reduced to some specified negligible value at a recommended V_{DS}. This gate voltage corresponds to the pinch-off voltage V_P of a JFET.

The foregoing discussion is applicable in principle also to the p-channel MOSFET. For such a device the signs of all currents and voltages in the volt-ampere characteristics of Fig. 10-13 must be reversed.

Comparison of p- with n-Channel FETs The p-channel enhancement FET, shown in Fig. 10-9, is very popular in MOS systems because it is much easier to produce than the n-channel device. Most of the contaminants in MOS fabrication are mobile ions which are positively charged and are trapped in the oxide layer between gate and substrate. In an n-channel enhancement device the gate is normally positive with respect to the substrate and, hence, the positively charged contaminants collect along the interface between the SiO_2 and the silicon substrate. The positive charge from this layer of ions attracts free electrons in the channel which tends to make the transistor turn on prematurely. In p-channel devices the positive contaminant ions are pulled to the opposite side of the oxide layer (to the aluminum-SiO_2 interface) by the negative gate voltage and there they cannot affect the channel.

The hole mobility in silicon and at normal field intensities is approximately 500 cm²/V-s. On the other hand, electron mobility is about 1,300 cm²/V-s. Thus the p-channel device will have more than twice the ON resistance of an equivalent *n*-channel of the same geometry and under the same operating conditions. In other words, the *p*-channel device must have more than twice the area of the *n*-channel device to achieve the same resistance. Therefore *n*-channel MOS circuits can be smaller for the same complexity than *p*-channel devices. The higher packing density of the *n*-channel MOS also makes it faster in switching applications due to the smaller junction areas. The operating speed is limited primarily by the internal *RC* time constants, and the capacitance is directly proportional to the junction cross sections. For all the above reasons it is clear that *n*-channel MOS circuits are more desirable than *p*-channel circuits. However, the more extensive process control needed for *n*-channel fabrication makes them expensive and unable to compete economically with *p*-channel devices at this time (1972).

MOSFET Gate Protection Since the SiO_2 layer of the gate is extremely thin, it may easily be damaged by excessive voltage. An accumulation of charge on an open-circuited gate may result in a large enough field to punch through the dielectric. To prevent this damage some MOS devices are fabricated with a Zener diode between gate and substrate. In normal operation this diode is open and has no effect upon the circuit. However, if the voltage at the gate becomes excessive, then the diode breaks down and the gate potential is limited to a maximum value equal to the Zener voltage.

Circuit Symbols It is possible to bring out the connection to the substrate externally so as to have a tetrode device. Most MOSFETs, however, are triodes, with the substrate internally connected to the source. The circuit symbols used by several manufacturers are indicated in Fig. 10-14. Often the substrate lead is omitted from the symbol as in (*a*), and is then under-

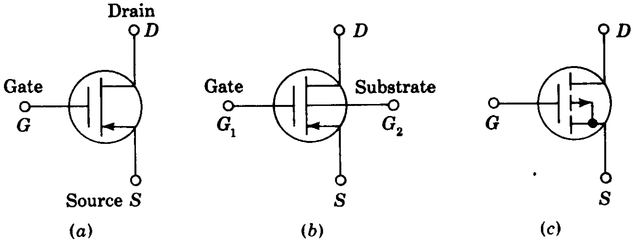
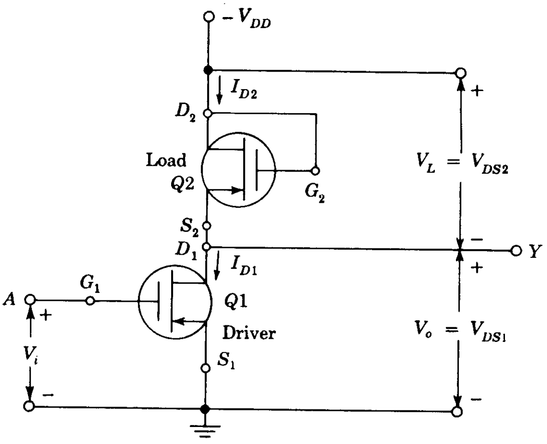


Fig. 10-14 Three circuit symbols for a *p*-channel MOSFET. (*a*) and (*b*) can be either depletion or enhancement types, whereas (c) represents specifically an enhancement device. In (*a*) the substrate is understood to be connected internally to the source. For an *n*-channel MOSFET the direction of the arrow is reversed.



(*a*)

A	Y
V_i	V_o
$-V_{DD}$	0
0	$-V_{DD}$

$Y = \bar{A}$

(*b*)

Fig. 10-15 (*a*) MOS inverter (NOT circuit). (*b*) The voltage truth table and Boolean expression.

stood to be connected to the source internally. For the enhancement-type MOSFET of Fig. 10-14*c*, G_2 is shown to be internally connected to *S*.

Small-signal MOSFET Circuit Model[11] If the small resistances of the source and drain regions are neglected, the small-signal equivalent circuit of the MOSFET between terminals G ($= G_1$), S, and D is identical with that given in Fig. 10-8 for the JFET. The transconductance g_m and the interelectrode capacitances have comparable values for the two types of devices. However, as noted in Table 10-1 on page 321, the drain resistance r_d of the MOSFET is very much smaller than that of the JFET. It should also be noted in Table 10-1 that the input resistance r_{gs} and the feedback resistance r_{gd} are very much larger for the MOSFET than for the JFET.

If the substrate terminal G_2 is not connected to the source, the model of Fig. 10-8 must be generalized as follows: Between node G_2 and S, a diode $D1$ is added to represent the *p-n* junction between the substrate and the source. Similarly, a second diode $D2$ is included between G_2 and D to account for the *p-n* junction formed by the substrate and the drain.

10-6 DIGITAL MOSFET CIRCUITS[12]

The most common applications of MOS devices are digital, such as logic gates (discussed in this section) and registers, or memory arrays (Chap. 17). Because of the gate-to-drain and gate-to-source and substrate parasitic capacitances, MOSFET circuits are slower than corresponding bipolar circuits.

However, the lower power dissipation and higher density of fabrication make MOS devices attractive and economical for many low-speed applications.

Inverter MOSFET digital circuits consist *entirely* of FETs and no other devices such as diodes, resistors, or capacitors (except for parasitic capacitances). For example, consider the MOSFET inverter of Fig. 10-15*a*. Device $Q1$ is the *driver FET*, whereas $Q2$ acts as its load resistance and is called the *load FET*. The nonlinear character of the load is brought into evidence as follows: Since the gate is tied to the drain, $V_{GS2} = V_{DS2}$. The drain characteristics of Fig. 10-10 are reproduced in Fig. 10-16*a*, and the shaded curve represents the locus of the points $V_{GS2} = V_{DS2} = V_L$. This curve also gives I_{D2} versus V_L (for $V_{GS2} = V_{DS2}$), and its slope gives the incremental load conductance g_L of $Q2$ as a load. Clearly, the load resistance is nonlinear. Note that $Q2$ is always conducting, (for $|V_{DS2}| > |V_T|$), regardless of whether $Q1$ is ON or OFF.

An analytical expression for the load curve is given by Eq. (10-8) with $V_{GS} = V_{DS} = V_L$ and with V_P replaced by the threshold voltage $V_{GST} = V_T$.

$$I_{DS} = I_{DSS}\left(1 - \frac{V_{DS}}{V_T}\right)^2 \qquad \text{for } |V_{DS}| \geq |V_T| \tag{10-18}$$

and we see that this is a quadratic, rather than a linear, relationship. From Eq. (10-18) we find (Prob. 10-9) that the load conductance is equal to the transconductance of the FET, $g_L = g_m$. The same result is obtained in Sec. 10-7.

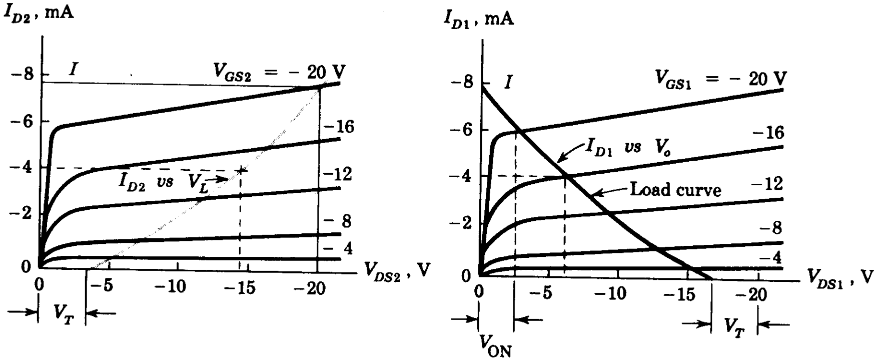
The incremental resistance is not a very useful parameter when considering large-signal (ON-OFF) digital operation. It is necessary to draw the *load curve* (corresponding to a *load line* with a constant resistance) on the volt-ampere characteristics of the driver FET $Q1$. The *load curve* is a plot of

$$I_D = I_{D1} \qquad \text{versus} \qquad V_{DS1} = V_o = -V_{DD} - V_L = -20 - V_{DS2}$$

where we have assumed a 20-V power supply. For a given value of $I_{D2} = I_{D1}$, we find $V_{DS2} = V_L$ from the shaded curve in Fig. 10-16*a* and then plot the locus of the values I_{D1} versus $V_o = V_{DS1}$ in Fig. 10-16*b*. For example, from Fig. 10-16*a* for $I_{D2} = 4$ mA, we find $V_{DS2} = -14$ V. Hence $I_{D1} = 4$ mA is located at $V_{DS1} = -20 + 14 = -6$ V in Fig. 10-16*b*.

We now confirm that the circuit of Fig. 10-15 is an inverter, or NOT circuit. Let us assume *negative* logic (Sec. 6-1) with the 1, or low state, given by $V(1) \approx -V_{DD} = -20$ V and the 0, or high state, given by $V(0) \approx 0$. If $V_i = V_{GS1} = -20$ V, then from Fig. 10-16*b*, $V_o = V_{\text{ON}} \approx -2$ V. Hence, $V_i = V(1)$ gives $V_o = V(0)$. Similarly from Fig. 10-16*b*, if $V_i = 0$ V, then $V_o = -V_{DD} - V_T = -17$ V for $V_T \approx -3$ V. Hence, $V_i = V(0)$ gives $V_o = V(1)$, thus confirming the truth table of Fig. 10-15*b*.

We shall simplify the remainder of the discussion in this section by assuming that $|V_{\text{ON}}|$ and $|V_T|$ are small compared with $|V_{DD}|$ and shall take $V_{\text{ON}} = 0$ and $V_T = 0$. Hence, to a first approximation the load FET may be con-



(a) (b)

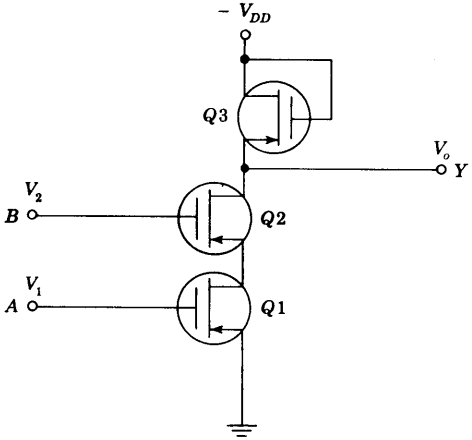
Fig. 10-16 (a) Q2 acts as a load I_{D2} versus $V_L = V_{DS2}$ for the driver $Q1$. (b) The load curve I_{D1} versus $V_o = V_{DS1}$.

sidered to be a constant resistance R_L and may be represented by a load line passing through $I_D = 0$, $V_{DS} = -V_{DD}$ and $I_D = I$, $V_{DS} = 0$, where I is the drain current for $V_{DS} = V_{GS} = -V_{DD}$. In other words, $R_L = -V_{DD}/I$. Most MOSFETs are p-channel enhancement-type devices, and negative logic is used with $V(0) = 0$ and $V(1) = -V_{DD}$.

NAND Gate The operation of the negative NAND gate of Fig. 10-17 can be understood if we realize that if either input V_1 or V_2 is at 0 V (the 0 state), the corresponding FET is OFF and the current is zero. Hence the voltage drop across the load FET is zero and the output $V_o = -V_{DD}$ (the 1 state). If both V_1 and V_2 are in the 1 state ($V_1 = V_2 = -V_{DD}$), then both $Q1$ and $Q2$ are ON and the output is 0 V, or at the 0 state. These values are in agreement with the voltage truth table of Fig. 10-17b. If 1 is substituted for $-V_{DD}$ in Fig. 10-17b, then this logic agrees with the truth table for a NAND gate, given in Fig. 6-18. We note that only during one of the four possible input states is power delivered by the power supply.

NOR Gate The circuit of Fig. 10-18a is a negative NOR gate. When either one of the two inputs (or both) is at $-V_{DD}$, the corresponding FET is ON and the output is at 0 V. If both inputs are at 0 V, both transistors $Q1$ and $Q2$ are OFF and the output is at $-V_{DD}$. These values agree with the truth table of Fig. 10-18b. Note that power is drawn from the supply during three of the four possible input states. Because of the high density of MOS devices on the same chip, it is important to minimize power consumption in LSI MOSFET systems (Sec. 17-17).

The circuit of Fig. 10-17 may be considered to be a *positive* NOR gate, and that of Fig. 10-18 to be a *positive* NAND gate (Sec. 6-9). These MOSFET circuits are examples of direct-coupled transistor logic (DCTL), mentioned



(a)

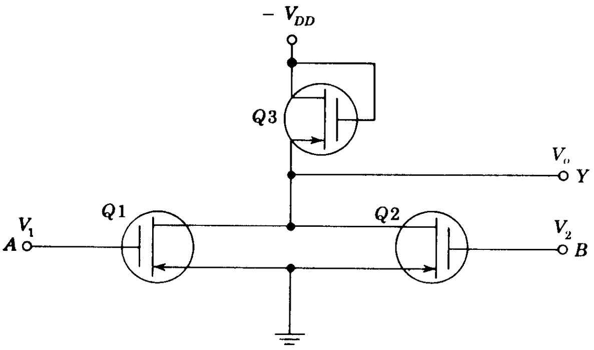
A	B	Y
V_1	V_2	V_0
0	0	$-V_{DD}$
0	$-V_{DD}$	$-V_{DD}$
$-V_{DD}$	0	$-V_{DD}$
$-V_{DD}$	$-V_{DD}$	0

$Y = \overline{AB}$

(b)

Fig. 10-17 (a) MOSFET (negative) NAND gate and (b) voltage truth table and Boolean expression. (Remember that 0 V is the zero state and $-V_{DD}$ is the 1 state.)

in Sec. 6-14. However, MOSFET DCTL circuits have none of the disadvantages (such as base-current "hogging") of bipolar DCTL gates. A FLIP-FLOP constructed from MOSFETs is indicated in Prob. 10-11. An AND (OR) gate is obtained by cascading a NAND (NOR) gate with a NOT gate. Typically, a three-input NAND gate uses about 16 mils^2 of chip area, whereas a single bipolar junction transistor may need about five times this area.



(a)

A	B	Y
V_1	V_2	V_o
0	0	$-V_{DD}$
0	$-V_{DD}$	0
$-V_{DD}$	0	0
$-V_{DD}$	$-V_{DD}$	0

$Y = \overline{A + B}$

(b)

Fig. 10-18 (a) MOSFET (negative) NOR gate and (b) voltage truth table and Boolean equation.

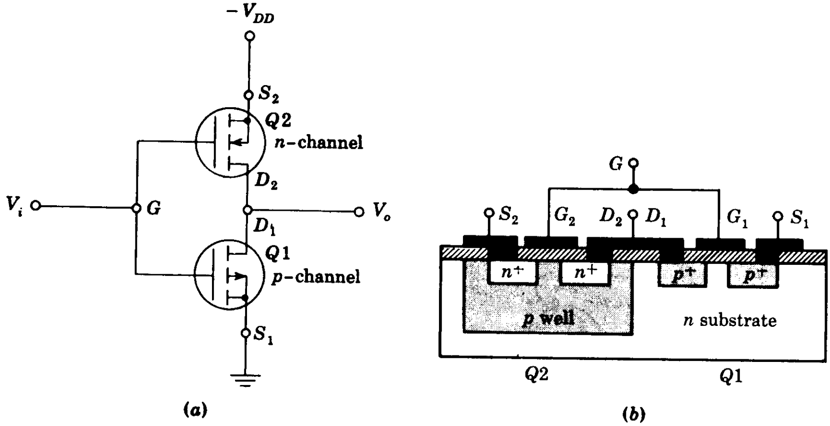


Fig. 10-19 (a) Complementary MOS inverter. (b) Cross section of complementary MOSFETS. Note that the p-type well is diffused into the n-type substrate and that the n-channel MOS $Q2$ is formed in this region.

Complementary MOS (CMOS)[12] It is possible to reduce the power dissipation to very small (50 nW) levels by using complementary p-channel and n-channel enhancement MOS devices on the same chip. The basic complementary MOS inverter circuit is shown in Fig. 10-19. Transistor $Q1$ is the p-channel unit, and transistor $Q2$ is n-channel. The two devices are in series, with their drains tied together and their gates also connected together. The logic swing gate voltage V_i varies from 0 V to the power supply $-V_{DD}$. When $V_i = -V_{DD}$ (logic 1) transistor $Q1$ is turned ON (but draws no appreciable steady-state current) and $Q2$ is turned OFF, the output V_o is then at 0 V (logic 0), and inversion has been accomplished. When zero voltage (logic 0) is applied at the input, the n-channel $Q2$ is turned ON (at no steady-state current) and $Q1$ is turned OFF. Thus the output is at $-V_{DD}$ (logic 1). In either logic state, $Q1$ or $Q2$ is OFF and the quiescent power dissipation for this simple inverter is the product of the OFF leakage current and $-V_{DD}$.

More complicated digital CMOS circuits (NAND, NOR, and FLIP-FLOPS) can be formed by combining simple inverter circuits (Probs. 10-13 and 10-14).

The remainder of the chapter considers the FET under small-signal operation. We first discuss low-frequency gain, then methods of biasing the device in the linear range, and finally the high-frequency limitations of the FET.

10-7 THE LOW-FREQUENCY COMMON-SOURCE AND COMMON-DRAIN AMPLIFIERS

The common-source (CS) stage is indicated in Fig. 10-20a, and the common-drain (CD) configuration in Fig. 10-20b. The former is analogous to the

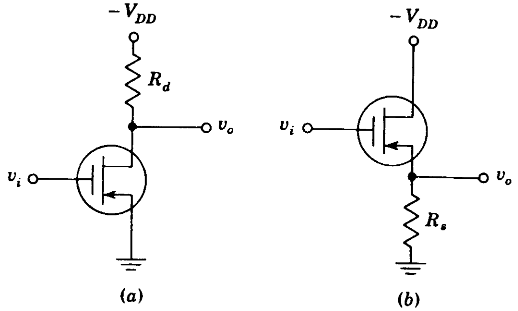


Fig. 10-20 (a) The CS and (b) the CD configurations.

bipolar transistor CE amplifier, and the latter to the CC stage. We shall analyze both of these circuits simultaneously by considering the generalized configuration in Fig. 10-21*a*. For the CS stage the output is v_{o1} taken at the drain and $R_s = 0$. For the CD stage the output is v_{o2} taken at the source and $R_d = 0$. The signal-source resistance is unimportant since it is in series with the gate, which draws negligible current. No biasing arrangements are indicated (Sec. 10-8), but it is assumed that the stage is properly biased for linear operation.

Replacing the FET by its low-frequency small-signal model of Fig. 10-8, the equivalent circuit of Fig. 10-21*b* is obtained. Applying KVL to the output circuit yields

$$i_dR_d + (i_d - g_mv_{gs})r_d + i_dR_s = 0 \tag{10-19}$$

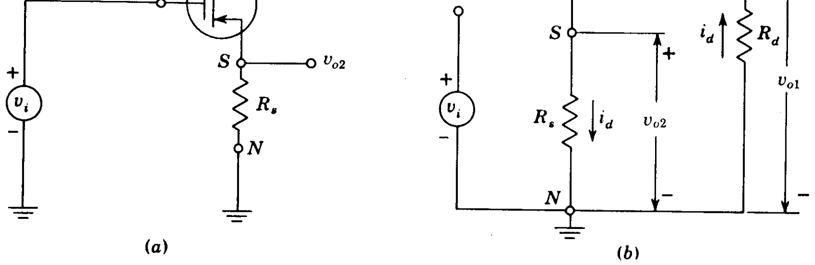


Fig. 10-21 (a) A generalized FET amplifier configuration. (b) The small-signal equivalent circuit.

From Fig. 10-21*b* the voltage from *G* to *S* is given by

$$v_{gs} = v_i - i_d R_s \tag{10-20}$$

Combining Eqs. (10-19) and (10-20) and remembering that $\mu = r_d g_m$ [Eq. (10-15)], we find

$$i_d = \frac{\mu v_i}{r_d + R_d + (\mu + 1)R_s} \tag{10-21}$$

The CS Amplifier with an Unbypassed Source Resistance Since $v_{o1} = -i_d R_d$, then

$$v_{o1} = \frac{-\mu v_i R_d}{r_d + R_d + (\mu + 1)R_s} \tag{10-22}$$

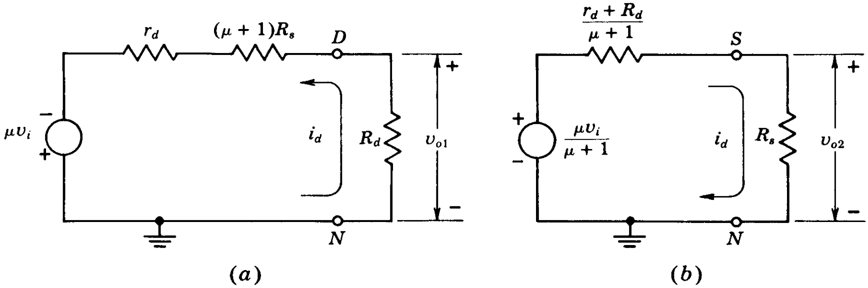
From Eq. (10-22) we obtain the Thévenin's equivalent circuit of Fig. 10-22*a* "looking into" the drain node (to ground). The open-circuit voltage is $-\mu v_i$, and the output resistance is $R_o = r_d + (\mu + 1)R_s$. The voltage gain is $A_V = v_{o1}/v_i$. The minus sign in Eq. (10-22) indicates that the output is 180° out of phase with the input. If R_s is bypassed with a large capacitance or if the source is grounded, the above equations are valid with $R_s = 0$. Under these circumstances,

$$A_V = \frac{v_{o1}}{v_i} = \frac{-\mu R_d}{r_d + R_d} = -g_m R_d' \tag{10-23}$$

where $\mu = r_d g_m$ [Eq. (10-15)] and $R_d' = R_d \| r_d$.

The CD Amplifier with a Drain Resistance Since $v_{o2} = i_d R_s$, then from Eq. (10-21)

$$v_{o2} = \frac{\mu v_i R_s}{r_d + R_d + (\mu + 1)R_s} = \frac{[\mu v_i/(\mu + 1)]R_s}{(r_d + R_d)/(\mu + 1) + R_s} \tag{10-24}$$



(*a*) (*b*)

Fig. 10-22 The equivalent circuits for the generalized amplifier of Fig. 10-21 "looking into" (*a*) the drain and (*b*) the source. Note that $\mu = r_d g_m$.

From Eq. (10-24) we obtain the Thévenin's equivalent circuit of Fig. 10-22*b* "looking into" the source node (to ground). The open-circuit voltage is $\mu v_i/(\mu + 1)$, and the output resistance is $R_o = (r_d + R_d)/(\mu + 1)$. The voltage gain is $A_V = v_{o2}/v_i$. Note that there is no phase shift between input and output. If $R_d = 0$ and if $(\mu + 1)R_s \gg r_d$, then $A_V \approx \mu/(\mu + 1) \approx 1$ for $\mu \gg 1$. A voltage gain of unity means that the output (at the source) follows the input (at the gate). Hence the CD configuration is called a *source follower* (analogous to the *emitter* follower for a bipolar junction transistor).

Note that the open-circuit voltage and the output impedance in either Fig. 10-22*a* or *b* are independent of the load (R_d in Fig. 10-22*a* and R_s in Fig. 10-22*b*). These restrictions must be satisfied if the networks in Fig. 10-22 are to represent the true Thévenin equivalents of the amplifier in Fig. 10-21.

For the source follower ($R_d = 0$) with $\mu \gg 1$, the output conductance is

$$g_o = \frac{1}{R_o} = \frac{\mu + 1}{r_d} \approx \frac{\mu}{r_d} = g_m \tag{10-25}$$

which agrees with the result obtained in Sec. 10-6 for the conductance looking into the source of a MOSFET with the gate at a constant voltage. In the discussion of diffused resistors in Sec. 7-8, it is indicated that 30 K is about the maximum resistance that can be fabricated. Larger values may be obtained by using the MOS structure as a load with gate connected to drain and tied to a fixed voltage such as $Q2$ in Fig. 10-15. By using a low g_m FET, a high value of effective resistance is obtained. For example, for $g_m = 10\ \mu\text{A/V}$, we obtain $R_o = 1/g_m = 100$ K. This value of effective resistance requires approximately 5 mil^2 of chip area compared with 300 mil^2 to yield a diffused 20-K resistance.

10-8 BIASING THE FET

The selection of an appropriate operating point (I_D, V_{GS}, V_{DS}) for an FET amplifier stage is determined by considerations similar to those given to transistors, as discussed in Chap. 9. These considerations are output-voltage swing, distortion, power dissipation, voltage gain, and drift of drain current. In most cases it is not possible to satisfy all desired specifications simultaneously. In this section we examine several biasing circuits for field-effect devices.

Source Self-bias The configuration shown in Fig. 10-23 can be used to bias junction FET devices or depletion-mode MOS transistors. For a specified drain current I_D, the corresponding gate-to-source voltage V_{GS} can be obtained applying either Eq. (10-8) or the plotted drain or transfer characteristics. Since the gate current (and, hence, the voltage drop across R_g) is negligible, the source resistance R_s can be found as the ratio of V_{GS} to the desired I_D.

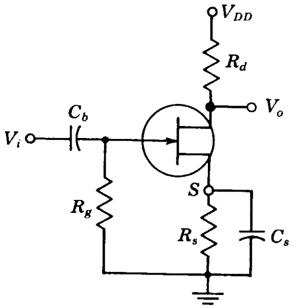


Fig. 10-23 Source self-bias circuit.

EXAMPLE The amplifier of Fig. 10-23 utilizes an n-channel FET for which $V_P = -2.0$ V and $I_{DSS} = 1.65$ mA. It is desired to bias the circuit at $I_D = 0.8$ mA, using $V_{DD} = 24$ V. Assume $r_d \gg R_d$. Find (*a*) V_{GS}, (*b*) g_m, (*c*) R_s, (*d*) R_d, such that the voltage gain is at least 20 dB, with R_s bypassed with a very large capacitance C_s.

Solution *a*. Using Eq. (10-8), we have $0.8 = 1.65(1 + V_{GS}/2.0)^2$. Solving, $V_{GS} = -0.62$ V.

b. Equation (10-17) now yields

$$g_{mo} = -\frac{2I_{DSS}}{V_P} = \frac{(2)(1.65)}{2} = 1.65 \text{ mA/V}$$

and from Eq. (10-16)

$$g_m = g_{mo}\left(1 - \frac{V_{GS}}{V_P}\right) = (1.65)\left(1 - \frac{0.62}{2.0}\right) = 1.14 \text{ mA/V}$$

c. $$R_s = -\frac{V_{GS}}{I_D} = \frac{0.62}{0.8} = 0.77 \text{ K} = 770\ \Omega$$

d. Since 20 dB corresponds to a voltage gain of 10, then from Eq. (10-23), with $r_d \gg R_d$, $|A_V| = g_m R_d \geq 10$, or $R_d \geq 10/1.14 = 8.76$ K.

Biasing against Device Variation FET manufacturers usually supply information on the maximum and minimum values of I_{DSS} and V_P at room temperature. They also supply data to correct these quantities for temperature variations. The transfer characteristics for a given type of n-channel FET may appear as in Fig. 10-24*a*, where the top and bottom curves are for extreme values of temperature and device variation. Assume that, on the basis of considerations previously discussed, it is necessary to bias the device at a drain current which will not drift outside of $I_D = I_A$ and $I_D = I_B$. Then the bias line $V_{GS} = -I_D R_s$ must intersect the transfer characteristics between the points A and B, as indicated in Fig. 10-24*a*. The slope of the bias line is determined by the source resistance R_s. For any transfer characteristic between the two extremes indicated, the current I_Q is such that $I_A < I_Q < I_B$, as desired.

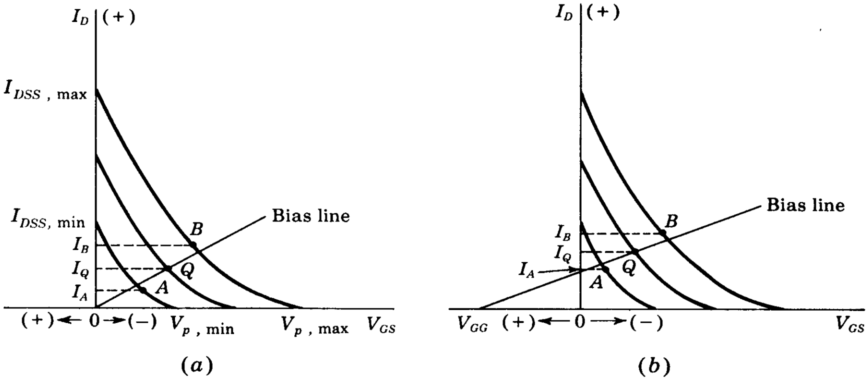


Fig. 10-24 Maximum and minimum transfer curves for an n-channel FET. The drain current must lie between I_A and I_B. The bias line can be drawn through the origin for the current limits indicated in (a), but this is not possible for the currents specified in (b).

Consider the situation indicated in Fig. 10-24b, where a line drawn to pass between points A and B does not pass through the origin. This bias line satisfies the equation

$$V_{GS} = V_{GG} - I_D R_s \tag{10-26}$$

Such a bias relationship may be obtained by adding a fixed bias to the gate in addition to the source self-bias, as indicated in Fig. 10-25a. A circuit

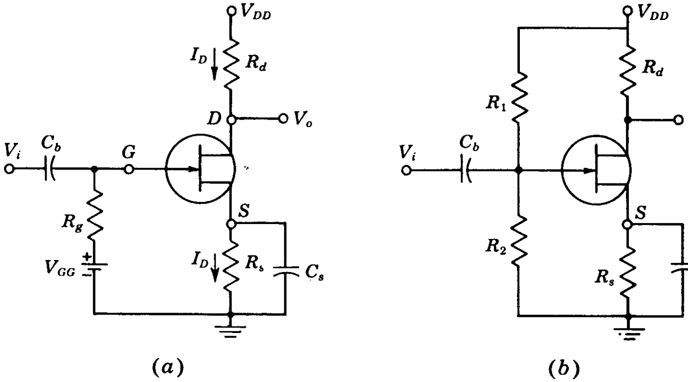


Fig. 10-25 (a) Biasing an FET with a fixed-bias V_{GG} in addition to self-bias through R_s. (b) A single power-supply configuration which is equivalent to the circuit in (a).

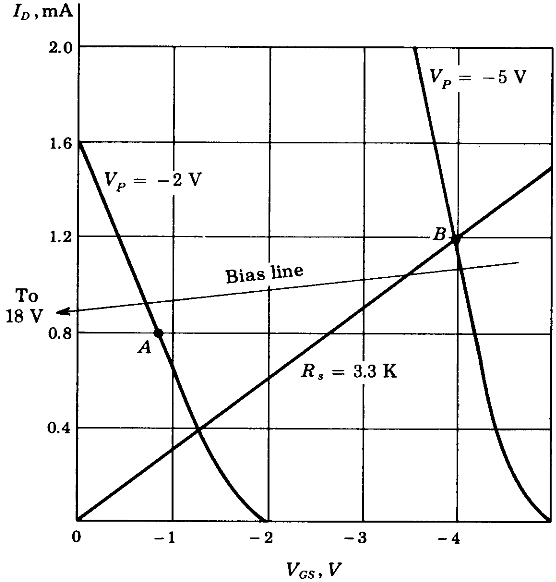


Fig. 10-26 Extreme transfer curves for the 2N3684 field-effect transistor. (Courtesy of Union Carbide Corporation.)

requiring only one power supply and which can satisfy Eq. (10-26) is shown in Fig. 10-25*b*. For this circuit

$$V_{GG} = \frac{R_2 V_{DD}}{R_1 + R_2} \qquad R_g = \frac{R_1 R_2}{R_1 + R_2}$$

We have assumed that the gate current is negligible. It is also possible for V_{GG} to fall in the reverse-biased region so that the line in Fig. 10-24*b* intersects the axis of abscissa to the right of the origin. Under these circumstances two separate supply voltages must be used.

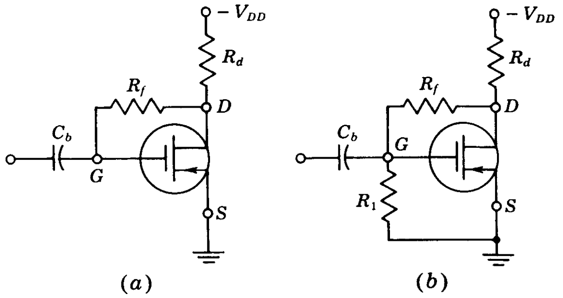
EXAMPLE FET 2N3684 is used in the circuit of Fig. 10-25*b*. For this *n*-channel device the manufacturer specifies $V_{P,\text{min}} = -2$ V, $V_{P,\text{max}} = -5$ V, $I_{DSS,\text{min}} = 1.6$ mA, and $I_{DSS,\text{max}} = 7.05$ mA. The extreme transfer curves are plotted in Fig. 10-26. It is desired to bias the circuit so that $I_{D,\text{min}} = 0.8$ mA $= I_A$ and $I_{D,\text{max}} = 1.2$ mA $= I_B$ for $V_{DD} = 24$ V. Find (*a*) V_{GG} and R_s, and (*b*) the range of possible values in I_D if $R_s = 3.3$ K and $V_{GG} = 0$.

Solution *a*. The bias line will lie between *A* and *B*, as indicated, if it is drawn to pass through the two points $V_{GS} = 0$, $I_D = 0.9$ mA, and $V_{GS} = -4$ V, $I_D = 1.1$ mA. The slope of this line determines R_s, or

$$R_s = \frac{4 - 0}{1.1 - 0.9} = 20 \text{ K}$$

Then, from the first point and Eq. (10-26), we find

$$V_{GG} = I_D R_s = (0.9)(20) = 18 \text{ V}$$



(a) (b)

Fig. 10-27 (a) Drain-to-gate bias circuit for enhancement-mode MOS transistors; (b) improved version of (a).

b. If $R_s = 3.3$ K, we see from the curves that $I_{D,\min} = 0.4$ mA and $I_{D,\max} = 1.2$ mA. The minimum current is far below the specified value of 0.8 mA.

Biasing the Enhancement MOSFET The self-bias technique of Fig. 10-23 cannot be used to establish an operating point for the enhancement-type MOSFET because the voltage drop across R_s is in a direction to reverse-bias the gate, and a forward gate bias is required. The circuit of Fig. 10-27*a* can be used, and for this case we have $V_{GS} = V_{DS}$, since no current flows through R_f. If for reasons of linearity in device operation or maximum output voltage it is desired that $V_{GS} \neq V_{DS}$, then the circuit of Fig. 10-27*b* is suitable. We note that $V_{GS} = [R_1/(R_1 + R_f)]V_{DS}$. Both circuits discussed here offer the advantages of dc stabilization through the feedback introduced with R_f. However, the input impedance is reduced because, by Miller's theorem (Sec. 8-11), R_f corresponds to an equivalent resistance $R_i = R_f/(1 - A_V)$ shunting the amplifier input.

Finally, note that the circuit of Fig. 10-25*b* is often used with the enhancement MOSFET. The dc stability introduced in Fig. 10-27 through the feedback resistor R_f is then missing, and is replaced by the dc feedback through R_s.

10-9 THE FET AS A VOLTAGE–VARIABLE RESISTOR[13] (VVR)

In most linear applications of field-effect transistors the device is operated in the constant-current portion of its output characteristics. We now consider FET transistor operation in the region before pinch-off, where V_{DS} is small. In this region the FET is useful as a voltage-controlled resistor; i.e., the drain-to-source resistance is controlled by the bias voltage V_{GS}. In such an application the FET is also referred to as a *voltage-variable resistor* (VVR), or *voltage-dependent resistor* (VDR).

Figure 10-28*a* shows the low-level bidirectional characteristics of an FET. The slope of these characteristics gives r_d as a function of V_{GS}. Figure 10-28*a* has been extended into the third quadrant to give an idea of device linearity around $V_{DS} = 0$.

In our treatment of the junction FET characteristics in Sec. 10-3, we derive Eq. (10-5), which gives the drain-to-source conductance $g_d = I_D/V_{DS}$ for small values of V_{DS}. From this equation we have

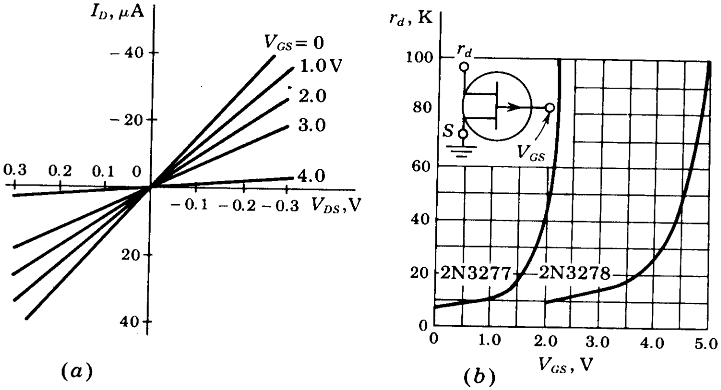
$$g_d = g_{do}\left[1 - \left(\frac{V_{GS}}{V_P}\right)^{\frac{1}{2}}\right] \tag{10-27}$$

where g_{do} is the value of the drain conductance when the bias is zero. Variation of r_d with V_{GS} is plotted in Fig. 10-28*b* for the 2N3277 and 2N3278 FETs. The variation of r_d with V_{GS} can be closely approximated by the empirical expression

$$r_d = \frac{r_o}{1 - KV_{GS}} \tag{10-28}$$

where r_o = drain resistance at zero gate bias
K = a constant, dependent upon FET type
V_{GS} = gate-to-source voltage

Applications of the VVR Since the FET operated as described above acts like a variable passive resistor, it finds applications in many areas where this property is useful. The VVR, for example, can be used to vary the voltage gain of a multistage amplifier A as the signal level is increased. This action is called *automatic gain control* (AGC). A typical arrangement is shown in Fig. 10-29. The signal is taken at a high-level point, rectified, and filtered to produce a dc voltage proportional to the output-signal level. This voltage is applied to the gate of $Q2$, thus causing the ac resistance between the drain and source to change, as shown in Fig. 10-28*b*. We thus may cause the gain of transistor $Q1$ to decrease as the output-signal level increases. The dc bias



(*a*) (*b*)

Fig. 10-28 (*a*) FET low-level drain characteristics for 2N3278. (*b*) Small-signal FET resistance variation with applied gate voltage. (Courtesy of Fairchild Semiconductor Company.)

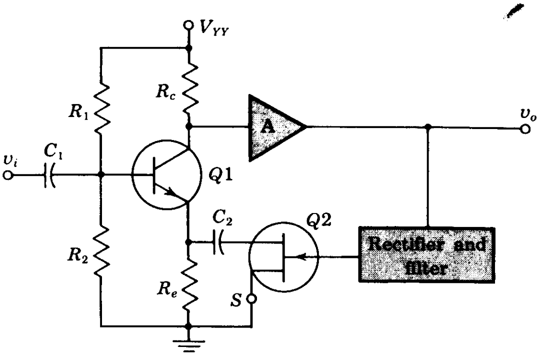


Fig. 10-29 AGC amplifier using the FET as a voltage-variable resistor.

conditions of $Q1$ are not affected by $Q2$ since $Q2$ is isolated from $Q1$ by means of capacitor C_2.

10-10 THE COMMON-SOURCE AMPLIFIER AT HIGH FREQUENCIES

The circuits discussed in this and the following section apply equally well to either JFETs or MOSFETs (except for the method of biasing). The low-frequency analysis of Sec. 10-8 is now modified to take into account the effect of the internal node capacitances.

Voltage Gain The circuit of Fig. 10-30*a* is the basic CS amplifier configuration. If the FET is replaced by the circuit model of Fig. 10-8*b*, we obtain the network in Fig. 10-30*b*. The output voltage V_o between D and S is easily found with the aid of the theorem of Sec. 8-7, namely, $V_o = IZ$, where I is the short-circuit current and Z is the impedance seen between the terminals.

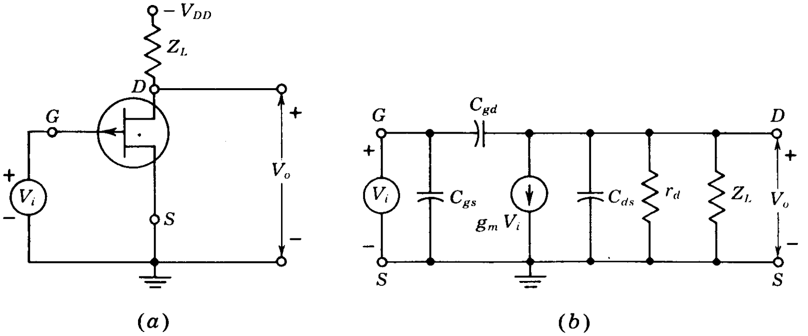


Fig. 10-30 (*a*) The common-source amplifier circuit; (*b*) small-signal equivalent circuit at high frequencies. (The biasing network is not indicated.)

To find Z, the independent generator V_i is (imagined) short-circuited, so that $V_i = 0$, and hence there is no current in the dependent generator $g_m V_i$. We then note that Z is the parallel combination of the impedances corresponding to Z_L, C_{ds}, r_d, and C_{gd}. Hence

$$Y = \frac{1}{Z} = Y_L + Y_{ds} + g_d + Y_{gd} \tag{10-29}$$

where $Y_L = 1/Z_L$ = admittance corresponding to Z_L
$Y_{ds} = j\omega C_{ds}$ = admittance corresponding to C_{ds}
$g_d = 1/r_d$ = conductance corresponding to r_d
$Y_{gd} = j\omega C_{gd}$ = admittance corresponding to C_{gd}

The current in the direction from D to S in a zero-resistance wire connecting the output terminals is

$$I = -g_m V_i + V_i Y_{gd} \tag{10-30}$$

The amplification A_V with the load Z_L in place is given by

$$A_V = \frac{V_o}{V_i} = \frac{IZ}{V_i} = \frac{I}{V_i Y} \tag{10-31}$$

or from Eqs. (10-29) and (10-30)

$$A_V = \frac{-g_m + Y_{gd}}{Y_L + Y_{ds} + g_d + Y_{gd}} \tag{10-32}$$

At low frequencies the FET capacitances can be neglected and hence

$$Y_{ds} = Y_{gd} = 0$$

Under these conditions Eq. (10-32) reduces to

$$A_V = \frac{-g_m}{Y_L + g_d} = \frac{-g_m r_d Z_L}{r_d + Z_L} = -g_m Z'_L \tag{10-33}$$

where $Z'_L \equiv Z_L \| r_d$. This equation agrees with Eq. (10-23), with Z_L replaced by R_d.

Input Admittance An inspection of Fig. 10-30*b* reveals that the gate circuit is not isolated from the drain circuit, but rather that they are connected by C_{gd}. From Miller's theorem (Sec. 8-11), this admittance may be replaced by $Y_{gd}(1 - K)$ between G and S, and by $Y_{gd}(1 - 1/K)$ between D and S, where $K = A_V$. Hence the input admittance is given by

$$Y_i = Y_{gs} + (1 - A_V)Y_{gd} \tag{10-34}$$

This expression indicates that for an FET to possess negligible input admittance over a wide range of frequencies, the gate-source and gate-drain capacitances must be negligible.

Input Capacitance (Miller Effect) Consider an FET with a drain-circuit resistance R_d. From the previous discussion it follows that within the audio-frequency range, the gain is given by the simple expression $A_V = -g_m R'_d$, where R'_d is $R_d \| r_d$. In this case, Eq. (10-34) becomes

$$\frac{Y_i}{j\omega} \equiv C_i = C_{gs} + (1 + g_m R'_d) C_{gd} \tag{10-35}$$

This increase in input capacitance C_i over the capacitance from gate to source is called the *Miller effect.*

This input capacitance is important in the operation of cascaded amplifiers. In such a system the output from one stage is used as the input to a second amplifier. Hence the input impedance of the second stage acts as a shunt across the output of the first stage and R_d is shunted by the capacitance C_i. Since the reactance of a capacitor decreases with increasing frequencies, the resultant output impedance of the first stage will be correspondingly low for the high frequencies. This will result in a decreasing gain at the higher frequencies.

EXAMPLE A MOSFET has a drain-circuit resistance R_d of 100 K and operates at 20 kHz. Calculate the voltage gain of this device as a single stage, and then as the first transistor in a cascaded amplifier consisting of two identical stages. The MOSFET parameters are $g_m = 1.6$ mA/V, $r_d = 44$ K, $C_{gs} = 3.0$ pF, $C_{ds} = 1.0$ pF, and $C_{gd} = 2.8$ pF.

Solution

$$Y_{gs} = j\omega C_{gs} = j2\pi \times 2 \times 10^4 \times 3.0 \times 10^{-12} = j3.76 \times 10^{-7}\ \mho$$

$$Y_{ds} = j\omega C_{ds} = j1.26 \times 10^{-7}\ \mho$$

$$Y_{gd} = j\omega C_{gd} = j3.52 \times 10^{-7}\ \mho$$

$$g_d = \frac{1}{r_d} = 2.27 \times 10^{-5}\ \mho$$

$$Y_d = \frac{1}{R_d} = 10^{-5}\ \mho$$

$$g_m = 1.60 \times 10^{-3}\ \mho$$

The gain of a one-stage amplifier is given by Eq. (10-32):

$$A_V = \frac{-g_m + Y_{gd}}{g_d + Y_d + Y_{ds} + Y_{gd}} = \frac{-1.60 \times 10^{-3} + j3.52 \times 10^{-7}}{3.27 \times 10^{-5} + j4.78 \times 10^{-7}}$$

It is seen that the j terms (arising from the interelectrode capacitances) are negligible in comparison with the real terms. If these are neglected, then $A_V = -48.8$. This value can be checked by using Eq. (10-23), which neglects interelectrode capacitances. Thus

$$A_V = \frac{-\mu R_d}{R_d + r_d} = \frac{-70 \times 100}{100 + 44} = -48.6 = -g_m R'_d$$

Since the gain is a real number, the input impedance consists of a capacitor whose value is given by Eq. (10-35):

$$C_i = C_{gs} + (1 + g_m R'_d) C_{gd} = 3.0 + (1 + 49)(1.0) = 53 \text{ pF}$$

Consider now a two-stage amplifier, each stage consisting of an FET operating as above. The gain of the second stage is that just calculated. However, in calculating the gain of the first stage, it must be remembered that *the input impedance of the second stage acts as a shunt on the output of the first stage.* Thus the drain load now consists of a 100-K resistance in parallel with 53 pF. To this must be added the capacitance from drain to source of the first stage since this is also in shunt with the drain load. Furthermore, any stray capacitances due to wiring should be taken into account. For example, for every 1-pF capacitance between the leads going to the drain and gate of the second stage, 50 pF is effectively added across the load resistor of the first stage! This clearly indicates the importance of making connections with very short direct leads in high-frequency amplifiers. Let it be assumed that the input capacitance, taking into account the various factors just discussed, is 200 pF. Then the load admittance is

$$Y_L = \frac{1}{R_d} + j\omega C_i = 10^{-5} + j2\pi \times 2 \times 10^4 \times 200 \times 10^{-12}$$

$$= 10^{-5} + j2.52 \times 10^{-5}\ \mho$$

The gain is given by Eq. (10-33):

$$A_V = \frac{-g_m}{g_d + Y_L} = \frac{-1.6 \times 10^{-3}}{2.27 \times 10^{-5} + 10^{-5} + j2.52 \times 10^{-5}}$$

$$= -30.7 + j23.7 = 38.8\underline{/143.3^\circ}$$

Thus the effect of the capacitances has been to reduce the magnitude of the amplification from 48.8 to 38.8 and to change the phase angle between the output and input from 180 to 143.3°.

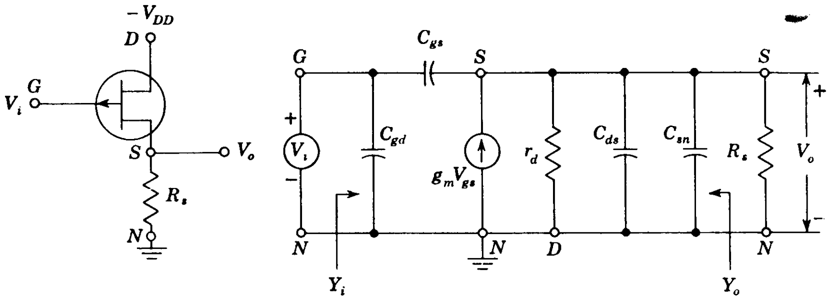
If the frequency were higher, the gain would be reduced still further. For example, this circuit would be useless as a video amplifier, say, to a few megahertz, since the gain would then be less than unity. This variation of gain with frequency is called *frequency distortion.* Cascaded amplifiers and frequency distortion are discussed in detail in Chap. 12.

Output Admittance For the common-source amplifier of Fig. 10-30 the output impedance is obtained by "looking into the drain" with the input voltage set equal to zero. If $V_i = 0$ in Fig. 10-30*b*, we see r_d, C_{ds}, and C_{gd} in parallel. Hence the output admittance with Z_L considered external to the amplifier is given by

$$Y_o = g_d + Y_{ds} + Y_{gd} \tag{10-36}$$

10-11 THE COMMON-DRAIN AMPLIFIER AT HIGH FREQUENCIES

The source-follower configuration is given in Fig. 10-20*b*, with $R_d = 0$, and is repeated in Fig. 10-31*a*. Its equivalent circuit with the FET replaced by its high-frequency model of Fig. 10-8*b* is shown in Fig. 10-31*b*.



(a) (b)

Fig. 10-31 (a) The source-follower; (b) small-signal high-frequency equivalent circuit. (The biasing network is not indicated.)

Voltage Gain The output voltage V_o can be found from the product of the short-circuit current and the impedance between terminals S and N. We now find for the voltage gain

$$A_V = \frac{(g_m + j\omega C_{gs})R_s}{1 + (g_m + g_d + j\omega C_T)R_s} \tag{10-37}$$

$$C_T \equiv C_{gs} + C_{ds} + C_{sn} \tag{10-38}$$

where C_{sn} represents the capacitance from source to ground. At low frequencies the gain reduces to

$$A_V \approx \frac{g_mR_s}{1 + (g_m + g_d)R_s} \tag{10-39}$$

Note that the amplification is positive and has a value less than unity. If $g_mR_s \gg 1$, then $A_V \approx g_m/(g_m + g_d) = \mu/(\mu + 1)$.

Input Admittance The source follower offers the important advantage of lower input capacitance than the CS amplifier. The input admittance Y_i is obtained by applying Miller's theorem to C_{gs}. We find

$$Y_i = j\omega C_{gd} + j\omega C_{gs}(1 - A_V) \approx j\omega C_{gd} \tag{10-40}$$

because $A_V \approx 1$.

Output Admittance The output admittance Y_o, with R_s considered external to the amplifier, is given by

$$Y_o = g_m + g_d + j\omega C_T \tag{10-41}$$

where C_T is given by Eq. (10-38). At low frequencies the output resistance R_o is

$$R_o = \frac{1}{g_m + g_d} \approx \frac{1}{g_m} \tag{10-42}$$

since $g_m \gg g_d$. For $g_m = 2$ mA/V, then $R_o = 500\ \Omega$.

The source follower is used for the same applications as the emitter follower, those requiring high input impedance and low output impedance.

REFERENCES

1. Shockley, W.: A Unipolar Field-effect Transistor, *Proc. IRE*, vol. 40, pp. 1365–1376, November, 1952.

   Dacey, G. C., and I. M. Ross: The Field Effect Transistor, *Bell System Tech. J.*, vol. 34, pp. 1149–1189, November, 1955.

   Wallmark, J. T., and H. Johnson: "Field-effect Transistors," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966.

   Sevin, L. J.: "Field-effect Transistors," McGraw-Hill Book Company, New York, 1965.

2. Millman, J., and H. Taub: "Pulse, Digital, and Switching Waveforms," sec. 17-20, McGraw-Hill Book Company, New York, 1965.
3. Wallmark, J. T., and H. Johnson: "Field-effect Transistors," p. 115, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1966.
4. Sevin, L. J., Ref. 1, pp. 13–17.
5. Halladay, H. E., and A. Van der Ziel: DC Characteristics of Junction Gate Field-effect Transistors, *IEEE Trans. Electron. Devices*, vol. ED-13, no. 6, pp. 531–532, June, 1966.
6. Sevin, L. J., Ref. 1, p. 21.
7. Sevin, L. J., Ref. 1, p. 23.
8. Sevin, L. J., Ref. 1, p. 34.
9. Ref. 3, pp. 187–215.
10. Macdougall, J., and K. Manchester: Ion Implantation, *Electronics*, vol. 43, no. 13, no. 13, pp. 86–90, June 22, 1970.
11. Ref. 3, pp. 256–259.
12. Garrett, L.: Integrated-circuit Digital Logic Families, *Spectrum*, vol. 7, no. 12, pp. 30–42, December, 1970.
13. Bilotti, A.: Operation of a MOS Transistor as a Voltage Variable Resistor, *Proc. IEEE*, vol. 54, pp. 1093–1094, August, 1966.

REVIEW QUESTIONS

10-1 (*a*) Sketch the basic structure of an n-channel junction field-effect transistor. (*b*) Show the circuit symbol for the JFET.

10-2 (*a*) Draw a family of CS drain characteristics of an *n*-channel JFET. (*b*) Explain the shape of these curves qualitatively.

10-3 How does the FET behave (*a*) for small values of $|V_{DS}|$? (*b*) For large $|V_{DS}|$?

10-4 (*a*) Define the *pinch-off voltage* V_P. (*b*) Sketch the depletion region before and after pinch-off.

10-5 Sketch the geometry of a JFET in integrated form.

10-6 (*a*) How does the drain current vary with gate voltage in the saturation region? (*b*) How does the transconductance vary with drain current?

10-7 Define (*a*) *transconductance* g_m, (*b*) *drain resistance* r_d, and (*c*) *amplification factor* μ of an FET.

10-8 Give the order of magnitude of g_m, r_d, and μ for a MOSFET.

10-9 Show the small-signal model of an FET (*a*) at low frequencies and (*b*) at high frequencies.

10-10 (*a*) Sketch the cross section of a *p*-channel enhancement MOSFET. (*b*) Show two circuit symbols for this MOSFET.

10-11 For the MOSFET in Rev. 10-10 draw (*a*) the drain characteristics and (*b*) the transfer curve.

10-12 Repeat Rev. 10-10 for an *n*-channel depletion MOSFET.

10-13 (*a*) Draw the circuit of a MOSFET NOT circuit. (*b*) Explain how it functions as an inverter.

10-14 (*a*) Explain how a MOSFET is used as a load. (*b*) Obtain the volt-ampere characteristic of this load graphically.

10-15 Sketch a two-input NAND gate and verify that it satisfies the Boolean NAND equation.

10-16 Repeat Rev. 10-15 for a two-input NOR gate.

10-17 Sketch a CMOS inverter and explain its operation.

10-18 (*a*) Draw the circuit of an FET amplifier with a source resistance R_s and a drain resistance R_d. (*b*) What is the Thévenin's equivalent circuit looking into the drain at low frequencies?

10-19 Repeat Rev. 10-18 looking into the source.

10-20 (*a*) Sketch the circuit of a source-follower. At low frequencies what is (*b*) the maximum value of the voltage gain? (*c*) The order of magnitude of the output impedance?

10-21 (*a*) Sketch the circuit of a CS amplifier. (*b*) Derive the expression for the voltage gain at low frequencies. (*c*) What is the maximum value of A_V?

10-22 (*a*) Draw two biasing circuits for a JFET or a depletion-type MOSFET. (*b*) Explain under what circumstances each of these two arrangements should be used.

10-23 Draw two biasing circuits for an enhancement-type MOSFET.

10-24 (*a*) How is an FET used as a voltage-variable resistance? (*b*) Explain.

10-25 (*a*) Sketch the small-signal high-frequency circuit of a CS amplifier. (*b*) Derive the expression for the voltage gain.

10-26 (*a*) From the circuit of Rev. 10-25, derive the input admittance. (*b*) What is the expression for the input capacitance in the audio range?

10-27 What specific capacitance has the greatest effect on the high-frequency response of a cascade of FET amplifiers? Explain.

10-28 Repeat Rev. 10-25 for a source-follower circuit.

10-29 Repeat Rev. 10-26 for a CD amplifier.

16 / INTEGRATED CIRCUITS AS ANALOG SYSTEM BUILDING BLOCKS

Many analog systems (both linear and nonlinear) are constructed with the OP AMP or DIFF AMP as the basic building block. These IC's augmented by a few external discrete components, either singly or in combination, are used in the following *linear* systems: analog computers, voltage-to-current and current-to-voltage converters, amplifiers of various types (for example, dc instrumentation, tuned, and video amplifiers), voltage followers, active filters, and delay equalizers.

Among the *nonlinear* analog system configurations discussed in this chapter are the following: amplitude modulators, logarithmic amplifiers and analog multipliers, sample-and-hold circuits, comparators, and square-wave and triangle-waveform generators.

I. LINEAR ANALOG SYSTEMS

16-1 BASIC OPERATIONAL AMPLIFIER APPLICATIONS[1]

An OP AMP may be used to perform many mathematical operations. This feature accounts for the name *operational amplifier*. Some of the basic applications are given in this section. Consider the ideal OP AMP of Fig. 15-2*a*, which is repeated for convenience in Fig. 16-1*a*. Recalling (Sec. 15-1) that the equivalent circuit of Fig. 16-1*b* has a virtual ground (which takes no current), it follows that the voltage gain is given by

$$A_{Vf} = \frac{V_o}{V_s} = -\frac{Z'}{Z} \tag{16-1}$$

Based upon this equation we can readily obtain an *analog inverter*, a *scale changer*, a *phase shifter*, and an *adder*.

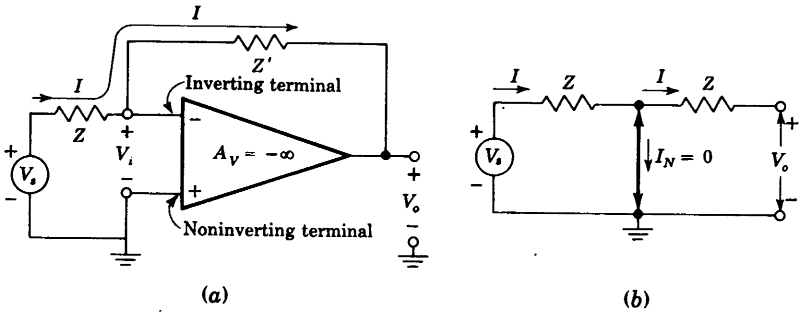


Fig. 16-1 (a) Inverting operational amplifier with voltage-shunt feedback. (b) Virtual ground in the OP AMP.

Sign Changer, or Inverter If $Z = Z'$ in Fig. 16-1, then $A_{Vf} = -1$, and the sign of the input signal has been changed. Hence such a circuit acts as a phase inverter. If two such amplifiers are connected in cascade, the output from the second stage equals the signal input without change of sign. Hence the outputs from the two stages are equal in magnitude but opposite in phase, and such a system is an excellent *paraphase amplifier.*

Scale Changer If the ratio $Z'/Z = k$, a real constant, then $A_{Vf} = -k$, and the scale has been multiplied by a factor $-k$. Usually, in such a case of multiplication by a constant, -1 or $-k$, Z and Z' are selected as precision resistors.

Phase Shifter Assume that Z and Z' are equal in magnitude but differ in angle. Then the operational amplifier shifts the phase of a sinusoidal input voltage while at the same time preserving its amplitude. Any phase shift from 0 to 360° (or ±180°) may be obtained.

Adder, or Summing Amplifier The arrangement of Fig. 16-2 may be used to obtain an output which is a linear combination of a number of input signals. Since a virtual ground exists at the OP AMP input, then

$$i = \frac{v_1}{R_1} + \frac{v_2}{R_2} + \cdots + \frac{v_n}{R_n}$$

and

$$v_o = -R'i = -\left(\frac{R'}{R_1} v_1 + \frac{R'}{R_2} v_2 + \cdots + \frac{R'}{R_n} v_n\right) \tag{16-2a}$$

If $R_1 = R_2 = \cdots = R_n$, then

$$v_o = -\frac{R'}{R_1}(v_1 + v_2 + \cdots + v_n) \tag{16-2b}$$

and the output is proportional to the sum of the inputs.

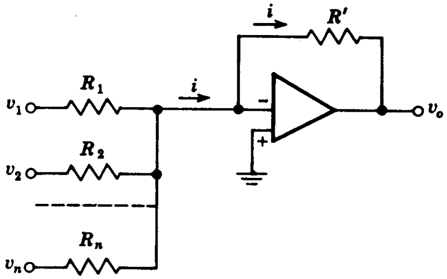


Fig. 16-2 Operational adder, or summing amplifier.

Many other methods may, of course, be used to combine signals. The present method has the advantage that it may be extended to a very large number of inputs requiring only one additional resistor for each additional input. The result depends, in the limiting case of large amplifier gain, only on the resistors involved, and because of the virtual ground, there is a minimum of interaction between input sources.

Voltage-to-current Converter Often it is desirable to convert a voltage signal to a proportional output current. This is required, for example, when we drive a deflection coil in a television tube. If the load impedance has neither side grounded (if it is floating), the simple circuit of Fig. 16-2 with R' replaced by the load impedance Z_L is an excellent *voltage-to-current converter.* For a single input $v_1 = v_s(t)$, the current in Z_L is

$$i_L = \frac{v_s(t)}{R} \tag{16-3}$$

Note that i is independent of the load Z_L, because of the virtual ground of the operational amplifier input. Since the same current flows through the signal source and the load, it is important that the signal source be capable of providing this load current. On the other hand, the amplifier of Fig. 16-3a requires

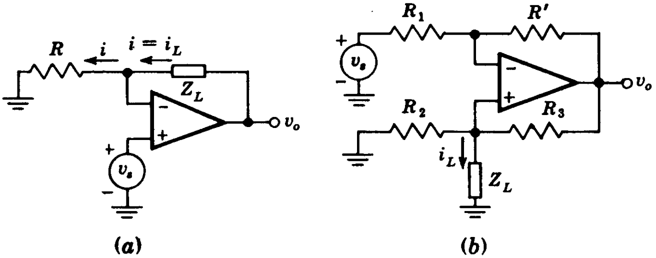


Fig. 16-3 Voltage-to-current converter for (a) a floating load and (b) a grounded load Z_L.

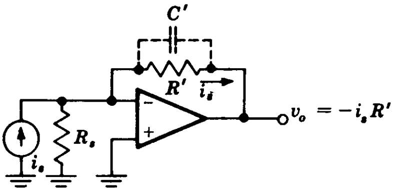


Fig. 16-4 Current-to-voltage converter.

very little current from the signal source due to the very large input resistance seen by the noninverting terminal.

If the load Z_L is grounded, the circuit of Fig. 16-3*b* can be used. In Prob. 16-7 we show that if $R_3/R_2 = R'/R_1$, then

$$i_L(t) = -\frac{v_s(t)}{R_2} \tag{16-4}$$

Current-to-voltage Converter Photocells and photomultiplier tubes give an output current which is independent of the load. The circuit in Fig. 16-4 shows an operational amplifier used as a current-to-voltage converter. Due to the virtual ground at the amplifier input, the current in R_s is zero and i_s flows through the feedback resistor R'. Thus the output voltage v_o is $v_o = -i_s R'$. It must be pointed out that the lower limit on current measurement with this circuit is set by the bias current of the inverting input. It is common to parallel R' with a capacitance C' to reduce high-frequency noise.

DC Voltage Follower The simple configuration of Fig. 16-5 approaches the ideal *voltage follower.* Because the two inputs are tied together (virtually), then $V_o = V_s$ and *the output follows the input.* The LM 102 (National Semiconductor Corporation) is specifically designed for voltage-follower usage and has very high input resistance (10,000 M), very low input current (~3 nA), and very low output resistance (~0 Ω).

16-2 DIFFERENTIAL DC AMPLIFIER[2]

The differential-input single-ended-output instrumentation amplifier is often used to amplify inputs from transducers which convert a physical parameter and its variations into an electric signal. Such transducers are strain-gauge bridges, thermocouples, etc. The circuit shown in Fig. 16-6 is very simple

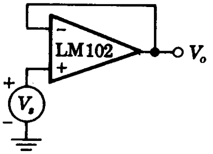
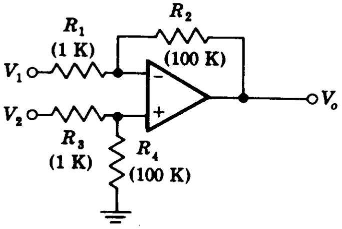


Fig. 16-5 A voltage follower, $V_o = V_s$.

Fig. 16-6 Differential amplifier using one OP AMP.



and uses only one OP AMP. In Prob. 16-8 we show that if $R_2/R_1 = R_4/R_3$, then

$$V_o = \frac{R_2}{R_1}(V_2 - V_1) \tag{16-5}$$

If the signals V_1 and V_2 have source resistances R_{s1} and R_{s2}, then these resistances add to R_1 and R_3, respectively. Note that the signal sources V_1 and V_2 are each loaded by only a 1-K resistance. If this is too heavy a load for the transducer, a voltage follower is used as a buffer amplifier for each signal. In other words, the circuit of Fig. 16-5 precedes each input in Fig. 16-6. This configuration of three IC OP AMPS is a very versatile, high-performance, low-cost dc amplifier system.

Bridge Amplifier A differential amplifier is often used to amplify the output from a transducer bridge, as shown in Fig. 16-7. Nominally, the four arms of the bridge have equal resistances R. However, one of the branches has a resistance which changes to $R + \Delta R$ with temperature or some other physical parameter. The goal of the measurement is to obtain the fractional change δ of the resistance value of the active arm, or $\delta = \Delta R/R$.

In Prob. 16-11 we find that for the circuit of Fig. 16-7, the output V_o is given by

$$V_o = -\frac{A_d V}{4}\frac{\delta}{1 + \delta/2} \tag{16-6}$$

For small changes in R ($\delta \ll 1$) Eq. (16-6) reduces to

$$V_o = -\frac{A_d V}{4}\delta \tag{16-7}$$

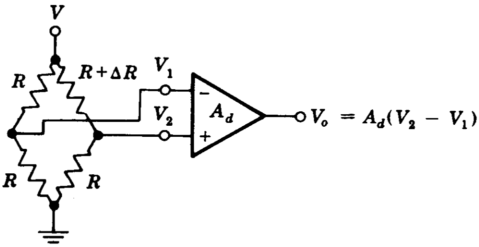


Fig. 16-7 Differential bridge amplifier.

16-3 STABLE AC-COUPLED AMPLIFIER

In some applications the need arises for the amplification of an ac signal, while any dc signal present must be blocked. A very simple and stable ac amplifier is shown in Fig. 16-8*a*, where capacitor C blocks the dc component of the input signal and together with the resistor R sets the low-frequency 3-dB response for the overall amplifier.

The output voltage V_o as a function of the complex variable s is found from the equivalent circuit of Fig. 16-8*b* (where the double-ended heavy arrow represents the virtual ground) to be

$$V_o = -IR' = -\frac{V_s}{R + 1/sC} R'$$

and

$$A_{Vf} = \frac{V_o}{V_s} = -\frac{R'}{R}\frac{s}{s + 1/RC} \tag{16-8}$$

From Eq. (16-8) we see that the low 3-dB frequency is

$$f_L = \frac{1}{2\pi RC} \tag{16-9}$$

The high-frequency response is determined by the frequency characteristics of the operational amplifier A_V and the amount of voltage-shunt feedback present (Sec. 14-5). The midband gain is, from Eq. (16-1), $A_{Vf} = -R'/R$.

AC Voltage Follower The ac voltage follower is used to provide impedance buffering, that is, to connect a signal source with high internal source resistance to a load of low impedance, which may even be capacitive. In Fig. 16-9 is shown a practical high-input impedance ac voltage follower using the LM 102 operational amplifier. We assume that C_1 and C_2 represent short circuits at all frequencies of operation of this circuit. Resistors R_1 and R_2 are used to provide RC coupling and allow a path for the dc input current into the noninverting terminal. In the absence of the bootstrapping capacitor C_2, the ac signal source would see an input resistance of only $R_1 + R_2 = 200$ K. Since

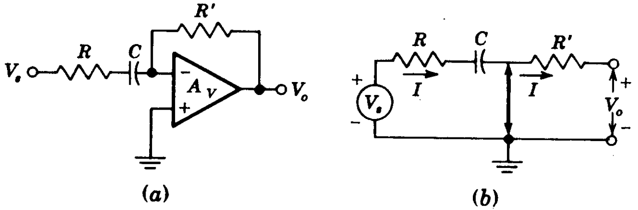
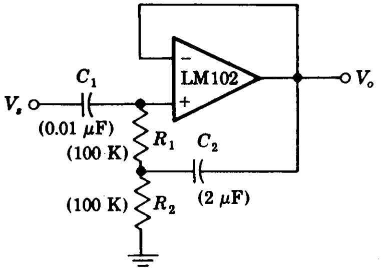


Fig. 16-8 (*a*) AC stable feedback amplifier. (*b*) Equivalent circuit when $|A_V| = \infty$.

Fig. 16-9 AC voltage follower. (Courtesy of National Semiconductor Corporation.)



the LM 102 is connected as a voltage follower, the voltage gain A_V between the output terminal and the noninverting terminal is very close to unity. Thus, from our discussion in Sec. 8-16, the input resistance the source sees becomes, approximately, $R_1/(1 - A_V)$, which is measured to be 12 M at 100 Hz and increases to 100 M at 1 kHz.

16-4 ANALOG INTEGRATION AND DIFFERENTIATION[1]

The analog integrator is very useful in many applications which require the generation or processing of analog signals. If, in Fig. 16-1, $Z = R$ and a capacitor C is used for Z', as in Fig. 16-10, we can show that the circuit performs the mathematical operation of integration. The input need not be sinusoidal, and hence is represented by the lowercase symbol $v = v(t)$. (The subscript s is now omitted, for simplicity.) In Fig. 16-10b, the double-headed arrow represents a virtual ground. Hence $i = v/R$, and

$$v_o = -\frac{1}{C}\int i\,dt = -\frac{1}{RC}\int v\,dt \tag{16-10}$$

The amplifier therefore provides an output voltage proportional to the integral of the input voltage.

If the input voltage is a constant, $v = V$, then the output will be a ramp, $v_o = -Vt/RC$. Such an integrator makes an excellent sweep circuit for a cathode-ray-tube oscilloscope, and is called a *Miller integrator*, or *Miller sweep*.[3]

DC Offset and Bias Current The input stage of the operational amplifier used in Fig. 16-10 is usually a DIFF AMP. The dc input offset voltage V_{io} appears across the amplifier input, and this voltage will be integrated and will appear at the output as a linearly increasing voltage. Part of the input bias current will also flow through the feedback capacitor, charging it and producing an additional linearly increasing voltage at the output. These two ramp voltages continue to increase until the amplifier reaches its saturation point.

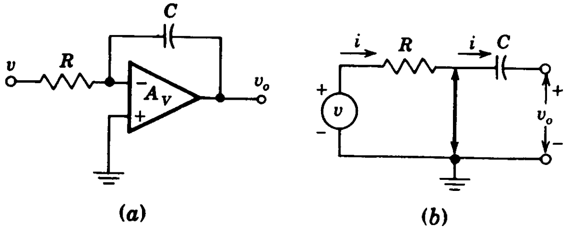


Fig. 16-10 (a) Operational integrator. (b) Equivalent circuit.

We see then that a limit is set on the feasible integration time by the above error components. The effect of the bias current can be minimized by increasing the feedback capacitor C while simultaneously decreasing the value of R for a given value of the time constant RC.

Finite Gain and Bandwidth The integrator supplies an output voltage proportional to the integral of the input voltage, provided the operational amplifier shown in Fig. 16-10a has infinite gain $|A_V| \to \infty$ and infinite bandwidth. The voltage gain as a function of the complex variable s is, from Eq. (16-1),

$$A_{Vf}(s) = \frac{V_o(s)}{V(s)} = -\frac{Z'}{Z} = -\frac{1}{RCs} \tag{16-11}$$

and it is clear that the ideal integrator has a pole at the origin.

Let us assume that in the absence of C the operational amplifier has a dominant pole at f_1, or $s_1 \equiv -2\pi f_1$. Hence its voltage gain A_v is approximated by

$$A_v = \frac{A_{vo}}{1 + j(f/f_1)} = \frac{A_{vo}}{1 - s/s_1} \tag{16-12}$$

If we further assume that $R_o = 0$ in Fig. 15-3, then $A_v = A_V$. Substituting Eq. (16-12) in Eq. (15-2) with $R_i = \infty$ and using $|A_{Vo}| \gg 1$, $|A_{Vo}|RC \gg 1/|s_1|$, we find

$$A_{Vf} = -\frac{s_1}{RC} A_{Vo} \frac{1}{(s + A_{Vo}s_1)(s - 1/RCA_{Vo})} \tag{16-13}$$

where A_{Vo} is a negative number and represents the low-frequency voltage gain of the operational amplifier.

The above transfer function has two poles on the negative real axis as compared with one pole at the origin for the ideal integrator. In Fig. 16-11 we show the Bode plots of Eqs. (16-11) to (16-13). We note that the response of the real integrator departs from the ideal at both low and high frequencies. At high frequencies the integrator performance is affected by the finite bandwidth $(-s_1/2\pi)$ of the operational amplifier, while at low frequencies the integration is limited by the finite gain of the OP AMP.

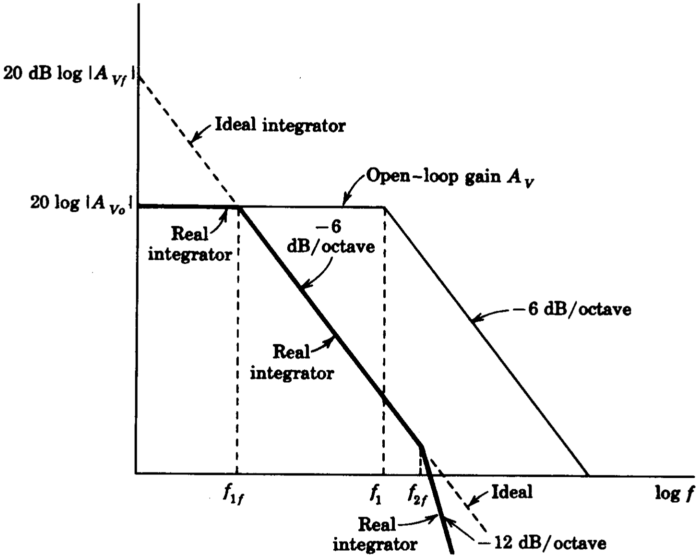
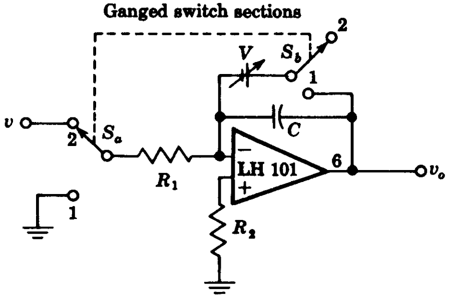


Fig. 16-11 Bode magnitude plots of open-loop OP AMP gain A_V, ideal integrator and real integrator. Note that $f_{1f} = 1/2\pi RC|A_{Vo}|$ and $f_{2f} = A_{Vo}s_1/2\pi$.

Practical Circuit A practical integrator must be provided with an external circuit to introduce initial conditions, as shown in Fig. 16-12. When switch S is in position 1, the input is zero and capacitor C is charged to the voltage V, setting an initial condition of $v_o = V$. When switch S is in position 2, the amplifier is connected as an integrator and its output will be V plus a constant times the time integral of the input voltage v. In using this circuit, care must be exercised to stabilize the amplifier and R_2 must be equal to R_1 to minimize the error due to bias current.

Fig. 16-12 Practical integrator circuit. For minimum offset error due to input bias current it is required that $R_1 = R_2$. (Courtesy of National Semiconductor Corporation.)



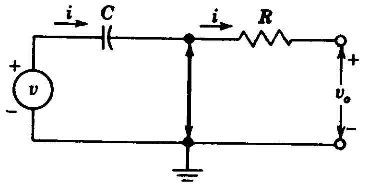


Fig. 16-13 Equivalent circuit of the operational differentiator.

Differentiator If Z is a capacitor C and if $Z' = R$, we see from the equivalent circuit of Fig. 16-13 that $i = C\,dv/dt$ and

$$v_o = -Ri = -RC\frac{dv}{dt} \tag{16-14}$$

Hence the output is proportional to the time derivative of the input. If the input signal is $v = \sin \omega t$, then the output will be $v_o = -RC\omega \cos \omega t$. Thus the magnitude of the output increases linearly with increasing frequency, and the differentiator circuit has high gain at high frequencies. This results in amplification of the high-frequency components of amplifier noise, and the noise output may completely obscure the differentiated signal.

The General Case In the important cases considered above, Z and Z' have been simple elements such as a single R or C. In general, they may be any series or parallel combinations of R, L, or C. Using the methods of Laplace transform analysis, Z and Z' can be written in their operational form as $Z(s)$ and $Z'(s)$, where s is the complex-frequency variable. In this notation the reactance of an inductor is written formally as Ls and that of a capacitor as $1/sC$. The current $I(s)$ is then $V(s)/Z(s)$, and the output is

$$V_o(s) = -\frac{Z'(s)}{Z(s)}V(s) \tag{16-15}$$

The amplifier thus solves this operational equation.

16-5 ELECTRONIC ANALOG COMPUTATION[1]

The OP AMP is the fundamental building block in an electronic analog computer. As an illustration, let us consider how to program the differential equation

$$\frac{d^2v}{dt^2} + K_1\frac{dv}{dt} + K_2v - v_1 = 0 \tag{16-16}$$

where v_1 is a given function of time, and K_1 and K_2 are real positive constants.

We begin by assuming that d^2v/dt^2 is available in the form of a voltage. Then, by means of an integrator, a voltage proportional to dv/dt is obtained. A second integrator gives a voltage proportional to v. Then an adder (and scale changer) gives $-K_1(dv/dt) - K_2v + v_1$. From the differential equation (16-16), this equals d^2v/dt^2, and hence the output of this summing amplifier is fed to the input terminal, where we had assumed that d^2v/dt^2 was available in the first place.

The procedure outlined above is carried out in Fig. 16-14. The voltage d^2v/dt^2 is assumed to be available at an input terminal. The integrator (1) has a time constant $RC = 1$ s, and hence its output at terminal 1 is $-dv/dt$. This voltage is fed to a similar integrator (2), and the voltage at terminal 2 is $+v$. The voltage at terminal 1 is fed to the inverter and scale changer (3), and its output at terminal 3 is $+K_1(dv/dt)$. This same operational amplifier (3) is used as an adder. Hence, if the given voltage $v_1(t)$ is also fed into it as shown, the output at terminal 3 also contains the term $-v_1$, or the net output is $+K_1(dv/dt) - v_1$. Scale changer–adder (4) is fed from terminals 2 and 3, and hence delivers a resultant voltage $-K_2v - K_1(dv/dt) + v_1$ at terminal 4. By Eq. (16-16) this must equal d^2v/dt^2, which is the voltage that was assumed to exist at the input terminal. Hence the computer is completed by connecting terminal 4 to the input terminal. (This last step is omitted from Fig. 16-14 for the sake of clarity of explanation.)

The specified initial conditions (the value of dv/dt and v at $t = 0$) must now be inserted into the computer. We note that the voltages at terminals 1 and 2 in Fig. 16-14 are proportional to dv/dt and v, respectively. Hence initial conditions are taken care of (as in Fig. 16-12) by applying the correct voltages V_1 and V_2 across the capacitors in integrators 1 and 2, respectively.

The solution is obtained by opening switches S_1 and S_2 and simultaneously

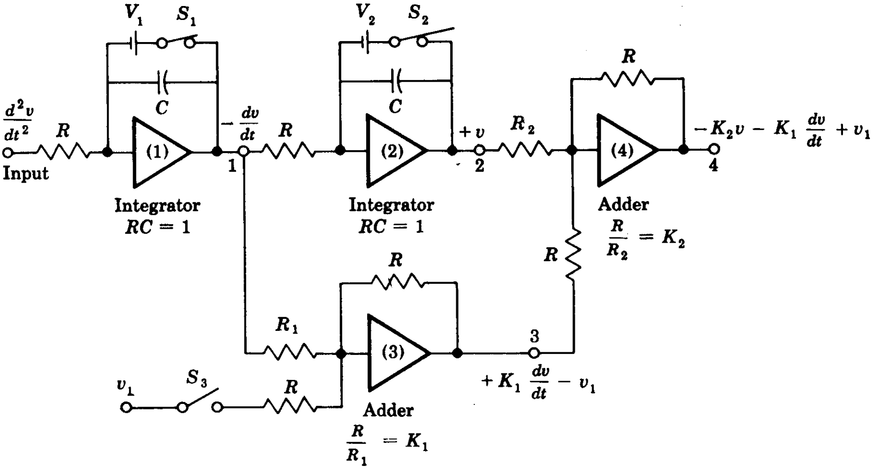


Fig. 16-14 A block diagram of an electronic analog computer. At $t = 0$, S_1 and S_2 are opened and S_3 is closed. Each OP AMP input is as in Fig. 16-12.

closing S_3 (by means of relays) at $t = 0$ and observing the waveform at terminal 2. If the derivative dv/dt is also desired, its waveform is available at terminal 1. The indicator may be a cathode-ray tube (with a triggered sweep) or a recorder or, for qualitative analysis with slowly varying quantities, a high-impedance voltmeter.

The solution of Eq. (16-16) can also be obtained with a computer which contains differentiators instead of integrators. However, integrators are almost invariably preferred over differentiators in analog-computer applications, for the following reasons: Since the gain of an integrator decreases with frequency whereas the gain of a differentiator increases nominally linearly with frequency, it is easier to stabilize the former than the latter with respect to spurious oscillations. As a result of its limited bandwidth, an integrator is less sensitive to noise voltages than a differentiator. Further, if the input waveform changes rapidly, the amplifier of a differentiator may overload. Finally, as a matter of practice, it is convenient to introduce initial conditions in an integrator.

16-6 ACTIVE FILTERS[4]

Consider the ideal low-pass-filter response shown in Fig. 16-15*a*. In this plot all signals within the band $0 \leq f \leq f_o$ are transmitted without loss, whereas inputs with frequencies $f > f_o$ give zero output. It is known[5] that such an ideal characteristic is unrealizable with physical elements, and thus it is

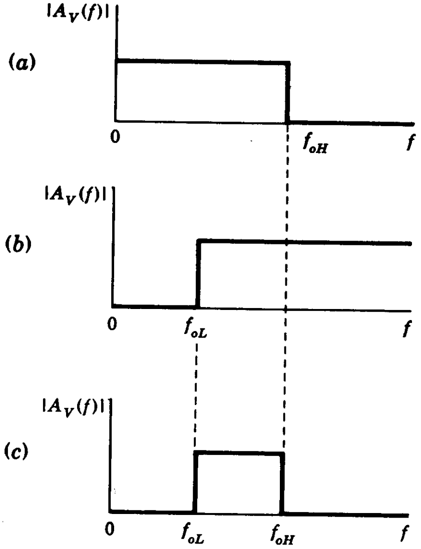


Fig. 16-15 Ideal filter characteristics. (*a*) Low-pass, (*b*) high-pass, and (*c*) bandpass.

necessary to approximate it. An approximation for an ideal low-pass filter is of the form

$$A_V(s) = \frac{1}{P_n(s)} \tag{16-17}$$

where $P_n(s)$ is a polynomial in the variable s with zeros in the left-hand plane. Active filters permit the realization of arbitrary left-hand poles for $A_V(s)$, using the operational amplifier as the active element and only resistors and capacitors for the passive elements.

Since commercially available OP AMPS have unity gain-bandwidth products as high as 100 MHz, it is possible to design active filters up to frequencies of several MHz. The limiting factor for full-power response at those high frequencies is the slewing rate (Sec. 15-6) of the operational amplifier. (Commercial integrated OP AMPS are available with slewing rates as high as 100 V/μs.)

Butterworth Filter[6] A common approximation of Eq. (16-17) uses the Butterworth polynomials $B_n(s)$, where

$$A_V(s) = \frac{A_{Vo}}{B_n(s)} \tag{16-18}$$

and with $s = j\omega$,

$$|A_V(s)|^2 = |A_V(s)|\,|A_V(-s)| = \frac{A_{Vo}{}^2}{1 + (\omega/\omega_o)^{2n}} \tag{16-19}$$

From Eqs. (16-18) and (16-19) we note that the magnitude of $B_n(\omega)$ is given by

$$|B_n(\omega)| = \sqrt{1 + \left(\frac{\omega}{\omega_o}\right)^{2n}} \tag{16-20}$$

The Butterworth response [Eq. (16-19)] for various values of n is plotted in Fig. 16-16. Note that the magnitude of A_V is down 3 dB at $\omega = \omega_o$ for all n. The larger the value of n, the more closely the curve approximates the ideal low-pass response of Fig. 16-15*a*.

If we normalize the frequency by assuming $\omega_o = 1$ rad/s, then Table 16-1 gives the Butterworth polynomials for n up to 8. Note that for n even, the polynomials are the products of quadratic forms, and for n odd, there is present the additional factor $s + 1$. The zeros of the normalized Butterworth polynomials are either -1 or complex conjugate and are found on the so-called *Butterworth circle* of unit radius shown in Fig. 16-17. The *damping factor* k is defined as one-half the coefficient of s in each quadratic factor in Table 16-1. For example, for $n = 4$, there are two damping factors, namely, $0.765/2 = 0.383$ and $1.848/2 = 0.924$. It turns out (Prob. 16-20) that k is given by

$$k = \cos\theta \tag{16-21}$$

where θ is as defined in Fig. 16-17*a* for n even and Fig. 16-17*b* for n odd.

TABLE 16-1 Normalized Butterworth polynominals

n	*Factors of polynomial* $P_n(s)$
1	$(s + 1)$
2	$(s^2 + 1.414s + 1)$
3	$(s + 1)(s^2 + s + 1)$
4	$(s^2 + 0.765s + 1)(s^2 + 1.848s + 1)$
5	$(s + 1)(s^2 + 0.618s + 1)(s^2 + 1.618s + 1)$
6	$(s^2 + 0.518s + 1)(s^2 + 1.414s + 1)(s^2 + 1.932s + 1)$
7	$(s + 1)(s^2 + 0.445s + 1)(s^2 + 1.247s + 1)(s^2 + 1.802s + 1)$
8	$(s^2 + 0.390s + 1)(s^2 + 1.111s + 1)(s^2 + 1.663s + 1)(s^2 + 1.962s + 1)$

From the table and Eq. (16-18) we see that the typical second-order Butterworth filter transfer function is of the form

$$\frac{A_V(s)}{A_{Vo}} = \frac{1}{(s/\omega_o)^2 + 2k(s/\omega_o) + 1} \qquad (16\text{-}22)$$

where $\omega_o = 2\pi f_o$ is the high-frequency 3-dB point. Similarly, the first-order filter is

$$\frac{A_V(s)}{A_{Vo}} = \frac{1}{s/\omega_o + 1} \qquad (16\text{-}23)$$

Practical Realization Consider the circuit shown in Fig. 16-18*a*, where the active element is an operational amplifier whose stable midband gain

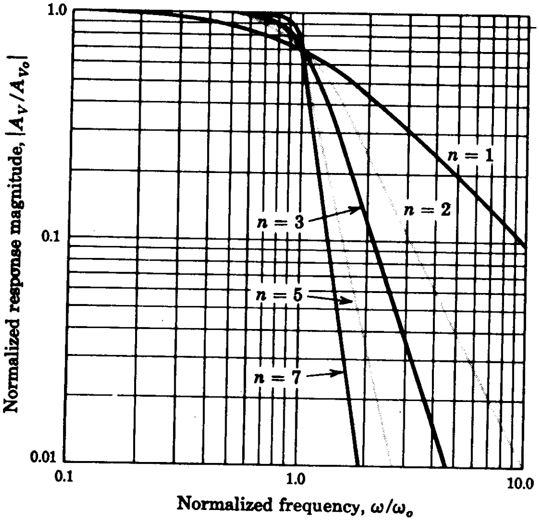


Fig. 16-16 Butterworth low-pass-filter frequency response.

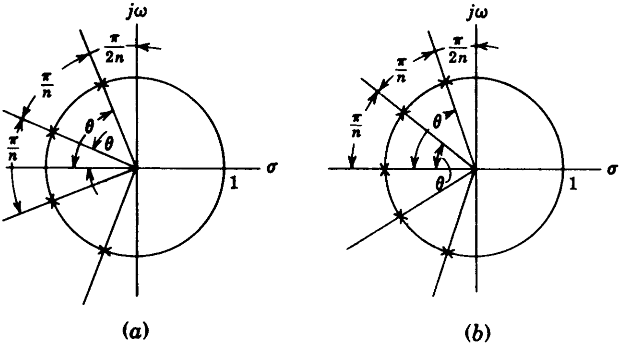


Fig. 16-17 The Butterworth circle for (a) n even and (b) n odd. Note that for n odd, one of the zeros is at $s = -1$.

$V_o/V_i = A_{Vo} = (R_1 + R_1')/R_1$ [Eq. (15-4)] is to be determined. We assume that the amplifier input current is zero, and we show in Prob. 16-25 that

$$A_V(s) = \frac{V_o}{V_s} = \frac{A_{Vo}Z_3Z_4}{Z_3(Z_1 + Z_2 + Z_3) + Z_1Z_2 + Z_1Z_4(1 - A_{Vo})} \tag{16-24}$$

If this network is to be a low-pass filter, then Z_1 and Z_2 are resistances and Z_3 and Z_4 are capacitances. Let us assume $Z_1 = Z_2 = R$ and $C_3 = C_4 = C$, as shown in Fig. 16-18b. The transfer function of this network takes the form

$$A_V(s) = A_{Vo}\frac{(1/RC)^2}{s^2 + \left(\dfrac{3 - A_{Vo}}{RC}\right)s + \left(\dfrac{1}{RC}\right)^2} \tag{16-25}$$

Comparing Eq. (16-25) with Eq. (16-22), we find

$$\omega_o = \frac{1}{RC} \tag{16-26}$$

and

$$2k = 3 - A_{Vo} \qquad \text{or} \qquad A_{Vo} = 3 - 2k \tag{16-27}$$

We are now in a position to synthesize even-order Butterworth filters by cascading prototypes of the form shown in Fig. 16-18b, using identical R's and C's and selecting the gain A_{Vo} of each operational amplifier to satisfy Eq. (16-27) and the damping factors from Table 16-1.

To realize odd-order filters, it is necessary to cascade the first-order filter of Eq. (16-23) with second-order sections such as indicated in Fig. 16-18b. The first-order prototype of Fig. 16-18c has the transfer function of Eq. (16-23) for arbitrary A_{Vo} provided that ω_o is given by Eq. (16-26). For example, a third-order Butterworth active filter consists of the circuit in Fig. 16-18b in cascade with the circuit of Fig. 16-18c, with R and C chosen so that $RC = 1/\omega_o$, with A_{Vo} in Fig. 16-18b selected to give $k = 0.5$ (Table 16-1, $n = 3$), and A_{Vo} in Fig. 16-18c chosen arbitrarily.

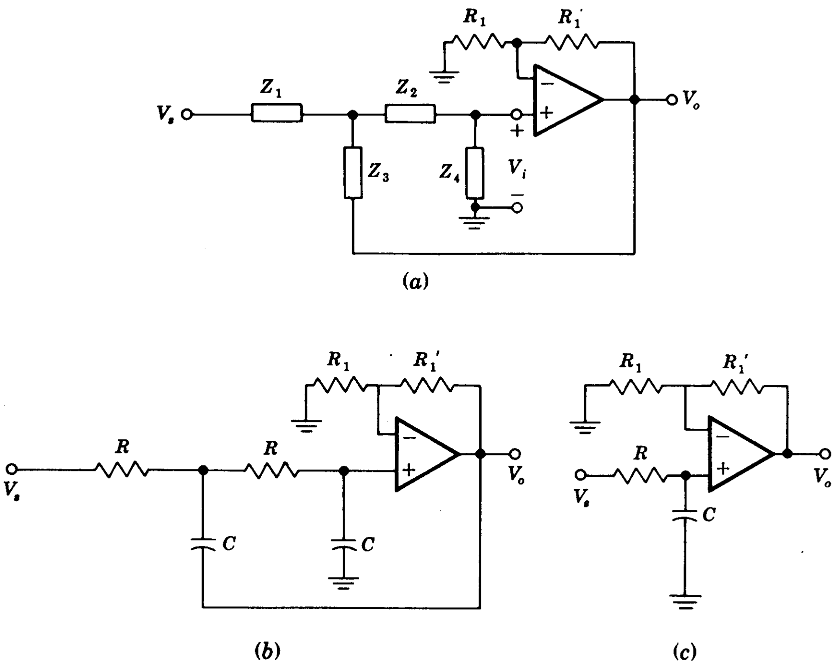

R_1
R_1'
Z_1
Z_2
V_s
V_o
Z_3
Z_4
V_i
+
−
(a)
R
C
(b)
(c)

Fig. 16-18 (a) Generalized active-filter prototype. (b) Second-order low-pass section. (c) First-order low-pass section.

EXAMPLE Design a fourth-order Butterworth low-pass filter with a cutoff frequency of 1 kHz.

Solution We cascade two second-order prototypes as shown in Fig. 16-19. For $n = 4$ we have from Table 16-1 and Eq. (16-27)

$$A_{V1} = 3 - 2k_1 = 3 - 0.765 = 2.235$$

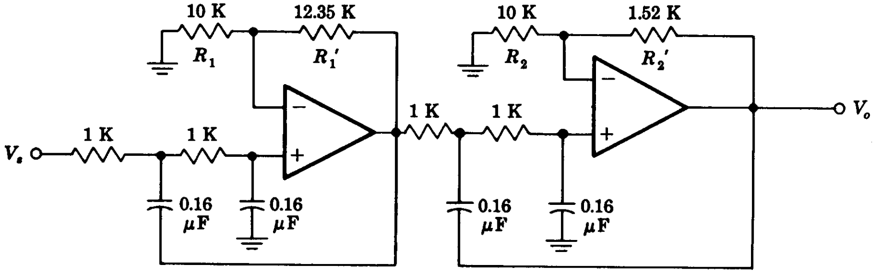

10 K
12.35 K
R_1
R_1'
10 K
1.52 K
R_2
R_2'
1 K
1 K
1 K
1 K
V_s
V_o
0.16 μF
0.16 μF
0.16 μF
0.16 μF

Fig. 16-19 Fourth-order Butterworth low-pass filter with $f_o = 1$ kHz.

and

$$A_{V2} = 3 - 2k_2 = 3 - 1.848 = 1.152$$

From Eq. (15-4), $A_{V1} = (R_1 + R_1')/R_1$. If we arbitrarily choose $R_1 = 10$ K, then for $A_{V1} = 2.235$, we find $R_1' = 12.35$ K, whereas for $A_{V2} = 1.152$, we find $R_2' = 1.520$ K and $R_2 = 10$ K. To satisfy the cutoff-frequency requirement, we have, from Eq. (16-26), $f_o = 1/2\pi RC$. We take $R = 1$ K and find $C = 0.16\ \mu$F. Figure 16-19 shows the complete fourth-order low-pass Butterworth filter.

High-pass Prototype An idealized high-pass-filter characteristic is indicated in Fig. 16-15*b*. The high-pass second-order filter is obtained from the low-pass second-order prototype of Eq. (16-22) by applying the transformation

$$\left.\frac{s}{\omega_o}\right|_{\text{low-pass}} \rightarrow \left.\frac{\omega_o}{s}\right|_{\text{high-pass}} \tag{16-28}$$

Thus, interchanging *R*'s and *C*'s in Fig. 16-18*b* results in a second-order high-pass active filter.

Bandpass Filter A second-order bandpass prototype is obtained by cascading a low-pass second-order section whose cutoff frequency is f_{oH} with a high-pass second-order section whose cutoff frequency is f_{oL}, provided $f_{oH} > f_{oL}$, as indicated in Fig. 16-15*c*.

Band-reject Filter Figure 16-20 shows that a band-reject filter is obtained by paralleling a high-pass section whose cutoff frequency is f_{oL} with a low-pass

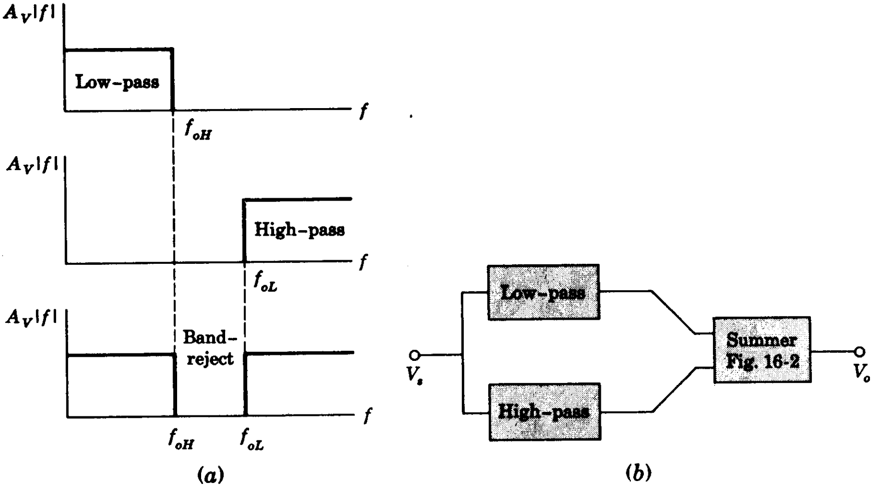


Fig. 16-20 (*a*) Ideal band-reject-filter frequency response. (*b*) Parallel combination of low-pass and high-pass filters results in a band-reject filter.

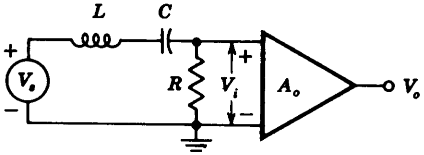


Fig. 16-21 A resonant circuit.

section whose cutoff frequency is f_{oH}. Note that for band-reject characteristics it is required that $f_{oH} < f_{oL}$.

16-7 ACTIVE RESONANT BANDPASS FILTERS[7]

The idealized bandpass filter of Fig. 16-15*c* has a constant response for $f_{oL} < f < f_{oH}$ and zero gain outside this range. An infinite number of Butterworth sections are required to obtain this filter response. A very simple approximation to a narrowband characteristic is obtained using a single *LC* resonant circuit. Such a bandpass filter has a response which peaks at some center frequency f_o and drops off with frequency on both sides of f_o. A basic prototype for a resonant filter is the second-order section shown in Fig. 16-21, whose transfer function we now derive.

If we assume that the amplifier provides a gain $A_o = V_o/V_i$ which is positive and constant for all frequencies, we find

$$A_V(j\omega) = \frac{V_o}{V_s} = \frac{V_oV_i}{V_iV_s} = \frac{RA_o}{R + j(\omega L - 1/\omega C)} \tag{16-29}$$

The *center*, or *resonant*, frequency $f_o = \omega_o/2\pi$ is defined as that frequency at which the inductance resonates with the capacitance; in other words, the inductive and capacitive reactances are equal (in magnitude), or

$$\omega_o^2 = \frac{1}{LC} \tag{16-30}$$

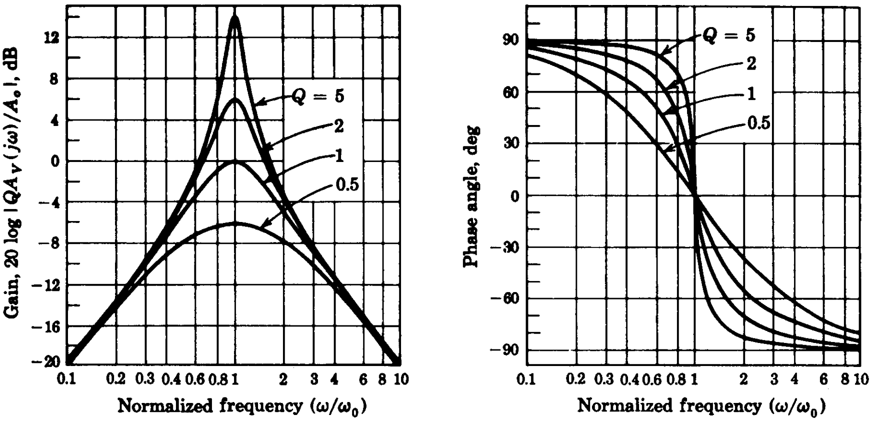
It is convenient to define the *quality factor Q* of this circuit by

$$Q \equiv \frac{\omega_o L}{R} = \frac{1}{\omega_o CR} = \frac{1}{R}\sqrt{\frac{L}{C}} \tag{16-31}$$

Substituting Eq. (16-31) in Eq. (16-29), we obtain the magnitude and phase of the transfer function

$$|A_V(j\omega)| = \frac{A_o}{\left[1 + Q^2\left(\frac{\omega}{\omega_0} - \frac{\omega_o}{\omega}\right)^2\right]^{\frac{1}{2}}} \tag{16-32}$$

$$\theta(\omega) = -\arctan Q\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right) \tag{16-33}$$



(a) *(b)*

Fig. 16-22 The bandpass characteristics of a tuned circuit. (*a*) Amplitude and (*b*) phase response.

Normalized Eqs. (16-32) and (16-33) are plotted in Fig. 16-22 for different values of the parameter Q.

Geometric Symmetry In the $|A_V(j\omega)|$ curves of Fig. 16-22 it is seen that, for every frequency $\omega' < \omega_o$, there exists a frequency $\omega'' > \omega_o$ for which $|A_V(j\omega)|$ has the same value. We now show that these frequencies have ω_o as their geometric mean; that is, $\omega_o^2 = \omega'\omega''$.

Setting $|A_V(j\omega')| = |A_V(j\omega'')|$, we obtain

$$\frac{\omega'}{\omega_o} - \frac{\omega_o}{\omega'} = -\left(\frac{\omega''}{\omega_o} - \frac{\omega_o}{\omega''}\right) \tag{16-34}$$

where the minus sign is required outside the parentheses because $\omega' < \omega_o < \omega''$. From Eq. (16-34) we find

$$\omega_o^2 = \omega'\omega'' \tag{16-35}$$

Bandwidth Let $\omega_1 < \omega_o$ and $\omega_2 > \omega_o$ be the two frequencies on either side of ω_o for which the gain drops by 3 dB from its value A_o at ω_o. Then the bandwidth is defined by

$$B \equiv \frac{\omega_2 - \omega_1}{2\pi} = \frac{1}{2\pi}\left(\omega_2 - \frac{\omega_o^2}{\omega_2}\right) \tag{16-36}$$

where use is made of Eq. (16-35). The frequency ω_2 is found by setting

$$\left|\frac{A_V(j\omega)}{A_o}\right| = \frac{1}{\sqrt{2}} \tag{16-37}$$

From Eq. (16-32) it follows that

$$Q\left(\frac{\omega_2}{\omega_o} - \frac{\omega_o}{\omega_2}\right) = 1 = \frac{Q}{\omega_o}\left(\omega_2 - \frac{\omega_o^2}{\omega_2}\right) \tag{16-38}$$

Comparing Eq. (16-36) with Eq. (16-38), we see that

$$B = \frac{1}{2\pi}\frac{\omega_o}{Q} = \frac{f_o}{Q} \tag{16-39}$$

The bandwidth is given by the center frequency divided by Q.

Substituting Eq. (16-31) in Eq. (16-39), we find an alternative expression for B, namely,

$$B = \frac{1}{2\pi}\frac{\omega_o R}{\omega_o L} = \frac{1}{2\pi}\frac{R}{L} \tag{16-40}$$

Active RC Bandpass Filter The general form for the second-order bandpass filter is obtained if we let $s = j\omega$ in Eq. (16-29).

$$A_V(s) = \frac{RA_o}{R + sL + 1/sC} = \frac{(R/L)A_o s}{s^2 + s(R/L) + 1/LC} \tag{16-41}$$

Substituting Eqs. (16-30) and (16-31) into (16-41) yields

$$A_V(s) = \frac{(\omega_o/Q)A_o s}{s^2 + (\omega_o/Q)s + \omega_o^2} \tag{16-42}$$

The transfer function of Eq. (16-42) obtained from the *RLC* circuit shown in Fig. 16-21 can be implemented with the multiple-feedback circuit of Fig. 16-23, which uses two capacitors, three resistors, and one OP AMP, but *no inductors*. If we assume that the OP AMP voltage gain is infinite, we show in Prob. 16-29 that

$$\frac{V_o(s)}{V_s} = \frac{s/R_1C_1}{s^2 + \dfrac{C_1 + C_2}{R_3C_1C_2}s + \dfrac{1}{R'R_3C_1C_2}} \tag{16-43}$$

where $R' = R_1 \| R_2$, or

$$R' \equiv \frac{R_1R_2}{R_1 + R_2} \tag{16-44}$$

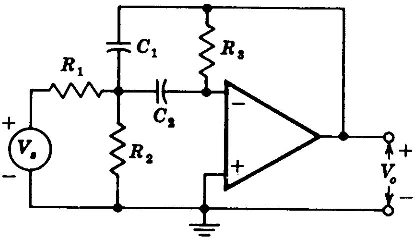


Fig. 16-23 An active resonant filter without an inductance.

Equating the corresponding coefficients in the three transfer functions of Eqs. (16-41), (16-42), and (16-43) yields

$$R_1C_1 = \frac{L}{RA_o} = \frac{Q}{\omega_o A_o} \tag{16-45}$$

$$R_3\frac{C_1C_2}{C_1 + C_2} = \frac{L}{R} = \frac{Q}{\omega_o} \tag{16-46}$$

$$R'R_3C_1C_2 = LC = \frac{1}{\omega_o{}^2} \tag{16-47}$$

Any real positive values for R_1, R', R_3, C_1, and C_2 which satisfy Eqs. (16-45) to (16-47) are acceptable for the design of the active bandpass filter. Since we have only three equations for the five parameters, two of these (say, C_1 and C_2) may be chosen arbitrarily.

EXAMPLE Design a second-order bandpass filter with a midband voltage gain $A_o = 50$ (34 dB), a center frequency $f_o = 160$ Hz, and a 3-dB bandwidth $B = 16$ Hz.

Solution From Eq. (16-39) we see that the required $Q = 160/16 = 10$. The center angular frequency is $\omega_o = 2\pi f_o = 2\pi \times 160 \approx 1{,}000$ rad/s. Assume $C_1 = C_2 = 0.1\ \mu$F. From Eq. (16-45)

$$R_1 = \frac{Q}{A_o\omega_oC_1} = \frac{10}{50 \times 10^3 \times 0.1 \times 10^{-6}}\ \Omega = 2\text{ K}$$

From Eq. (16-46)

$$R_3 = \frac{Q}{\omega_o\left(\dfrac{C_1C_2}{C_1 + C_2}\right)} = \frac{10}{1{,}000\left(\dfrac{0.1 \times 0.1}{0.2}\right) \times 10^{-6}}\ \Omega = 200\text{ K}$$

From Eq. (16-47)

$$R' = \frac{1}{\omega_o{}^2R_3C_1C_2} = \frac{1}{10^6 \times 2 \times 10^5 \times 10^{-14}} = 500\ \Omega$$

Finally, from Eq. (16-44)

$$R_2 = \frac{R_1R'}{R_1 - R'} = \frac{2{,}000 \times 500}{2{,}000 - 500} = 667\ \Omega$$

If the above specifications were to be met with the *RLC* circuit of Fig. 16-21, an unreasonably large value of inductance would be required (Prob. 16-32).

16-8 DELAY EQUALIZER

Signals such as digital data pulses transmitted over telephone wires suffer from delay distortion, discussed in Sec. 12-2. For the compensation of this distortion,

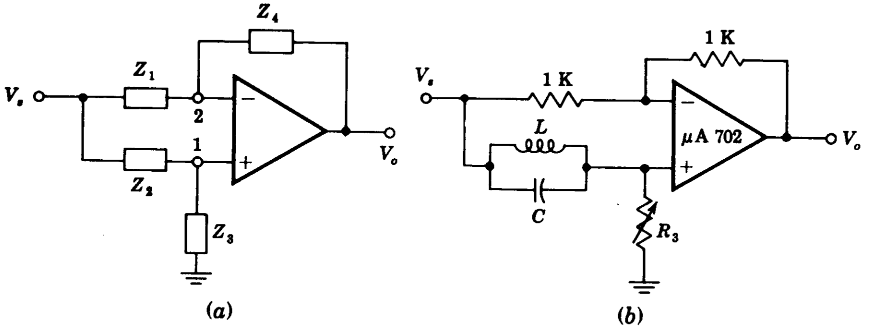


Fig. 16-24 (a) General form of delay equalizer. (b) Practical equalizer section using the μA702. (Courtesy of Fairchild Semiconductor, Inc.)

corrective networks known as *delay equalizers* are required. A delay equalizer is an all-pass network whose transfer function is of the form

$$A_V = \frac{V_o}{V_s} = \frac{R - jX}{R + jX} \tag{16-48}$$

We see from Eq. (16-48) that the amplitude of A_V is unity throughout the useful frequency range, and the delay D is given by the derivative of the phase of A_V with respect to frequency, or

$$D(\omega) = -2\frac{d}{d\omega}\left[\arctan\frac{X(\omega)}{R}\right] \tag{16-49}$$

A delay equalizer using an operational amplifier is shown in Fig. 16-24*a*. The transfer function for this configuration is found in Prob. 16-33 to be

$$A_V = \frac{Z_1Z_3 - Z_2Z_4}{Z_1(Z_3 + Z_2)} \tag{16-50}$$

For $Z_1 = Z_4 = R = 1$ K, $Z_3 = R_3$, and $Z_2 = jX$, Eq. (16-50) becomes

$$A_V = \frac{R_3 - jX}{R_3 + jX} \tag{16-51}$$

which is the desired all-pass characteristic of the delay equalizer. A practical delay equalizer is shown in Fig. 16-24*b* using the Fairchild μA702 OP AMP. The low offset voltage of this amplifier allows a larger number of sections to be directly coupled. This advantage is particularly significant when we consider the fact that in many applications eight or more sections in cascade are required to compensate for the delay distortion.

16-9 INTEGRATED CIRCUIT TUNED AMPLIFIER

The differential amplifier stage in monolithic integrated form (Fig. 16-25) is an excellent basic building block for the design of a tuned amplifier (including

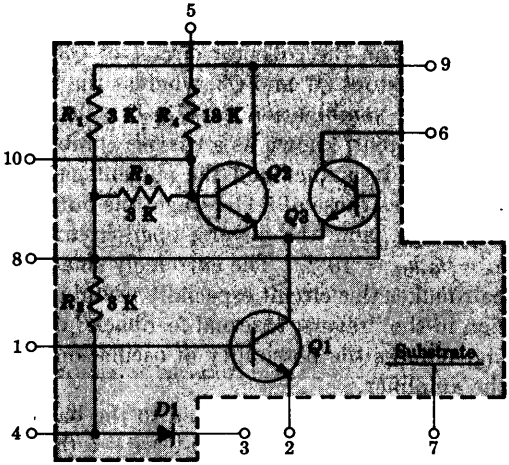

5
9
6
R_1 3 K
R_4 18 K
10
R_5
3 K
Q2
Q3
8
R_3 8 K
1
Q1
Substrate
D1
4
3
2
7

Fig. 16-25 The MC 1550 integrated circuit. (Courtesy of Motorola Semiconductor Inc.)

automatic gain control), an amplitude modulator, or a video amplifier. We now discuss these applications.

Operation of a Tuned Amplifier This circuit is designed to amplify a signal over a narrow band of frequencies centered at f_o. The *simplified* schematic diagram shown in Fig. 16-26 is used to explain the operation of this circuit. The external leads 1, 2, 3, . . . of the IC in this figure correspond to those in Fig. 16-25. The input signal is applied through the tuned trans-

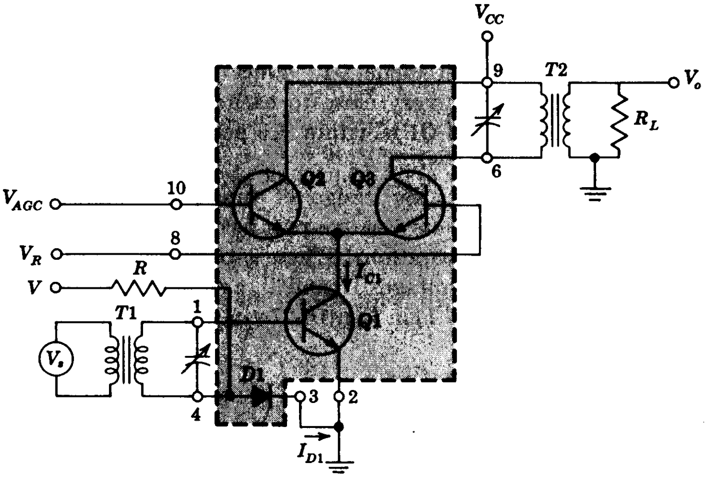

V_{CC}
9
T2
V_o
R_L
6
V_{AGC}
10
Q2
Q3
V_R
8
R
V
I_{C1}
T1
1
Q1
V_s
D1
3
2
4
I_{D1}

Fig. 16-26 Tuned amplifier consisting of the $Q1$-$Q3$ cascode, with the gain controlled by $Q2$.

former $T1$ to the base of $Q1$. The load R_L is applied across the tuned transformer $T2$ in the collector circuit of $Q3$. The amplification is performed by the transistors $Q1$ and $Q3$, whereas the magnitude of the gain is controlled by $Q2$. The combination of $Q1$-$Q3$ acts as a common-emitter common-base (CE-CB) pair, known as a *cascode* combination. In Prob. 8-39 we show that the input resistance and the current gain of a cascode circuit are essentially the same as those of a CE stage, the output resistance is the same as that of a CB stage, and the reverse-open-circuit voltage amplification is given by $h_r \approx h_{re}h_{rb} \approx 10^{-7}$. The extremely small value of h_r for the cascode transistor pair makes this circuit especially useful in tuned-amplifier design. The reduction in the "reverse internal feedback" of the compound device simplifies tuning, reduces the possibility of oscillation, and results in improved stability of the amplifier.

The voltage V_{AGC} applied to the base of $Q2$ is used to provide automatic gain control. From Fig. 15-9 we see that if V_{AGC} is at least 120 mV greater than V_R, $Q3$ is cut off and all the current of $Q1$ flows through $Q2$. Since $Q3$ is cut off, its transconductance is zero and the gain $A_V = V_o/V_s$ becomes zero. If V_{AGC} is less than V_R by more than 120 mV, $Q2$ is cut off and the collector current of $Q1$ flows through $Q3$, increasing the transconductance of $Q3$ and resulting in maximum voltage gain A_V.

An important advantage of this amplifier is its ability to vary the value of A_V by changing V_{AGC} without detuning the input circuit. This follows from the fact that variations in V_{AGC} cause changes in the division of the current between $Q2$ and $Q3$ without affecting significantly the collector current of $Q1$. Thus the input impedance of $Q1$ remains constant and the input circuit is not detuned.

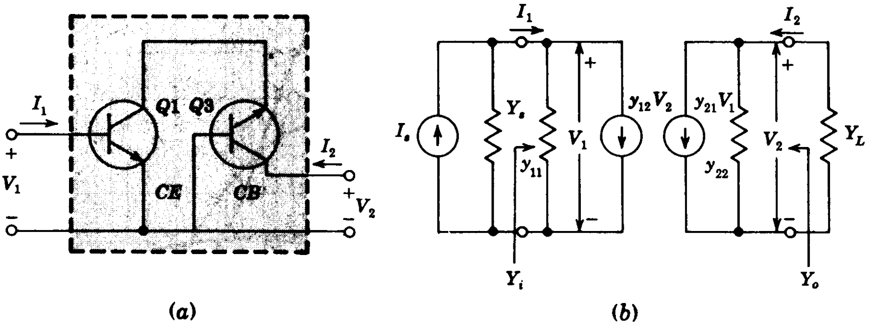
Biasing of this integrated amplifier is obtained using a technique similar to that discussed in Sec. 9-7. The voltage V and resistor R establish the dc current I_{D1} through the diode $D1$. Since the diode and transistor $Q1$ are on the same silicon chip, very close to each other, and with $V_{D1} = V_{BE1}$, the collector current I_{C1} of $Q1$ is within ± 5 percent of I_{D1}.

y-parameters In the design of tuned amplifiers, it is convenient to characterize the amplifiers as a two-port network and measure the y-parameters at the frequency of operation. These y-parameters are defined by choosing the input and output voltages V_1 and V_2 as independent variables and expressing the currents I_1 and I_2 in Fig. 16-27*a* in terms of these two voltages. Thus

$$I_1 = y_{11}V_1 + y_{12}V_2 \tag{16-52}$$

$$I_2 = y_{21}V_1 + y_{22}V_2 \tag{16-53}$$

where the I's and V's represent rms values of the small-signal currents and voltages. The circuit model satisfying these equations is indicated in Fig. 16-27*b*.



(a) (b)

Fig. 16-27 (a) Cascode pair. (b) y-parameter two-port model.

The y-parameter in Eqs. (16-52) and (16-53) and Fig. 16-27b are complex-valued functions of frequency which are defined as follows:

$$y_{11} \equiv G_{11} + jB_{11} \equiv \left.\frac{I_1}{V_1}\right|_{V_2=0} = \text{short-circuit input admittance}$$

$$y_{12} \equiv G_{12} + jB_{12} \equiv \left.\frac{I_1}{V_2}\right|_{V_1=0} = \text{short-circuit reverse transfer admittance}$$

$$y_{21} \equiv G_{21} - jB_{21} \equiv \left.\frac{I_2}{V_1}\right|_{V_2=0} = \text{short-circuit forward transfer admittance}$$

$$y_{22} \equiv G_{22} + jB_{22} \equiv \left.\frac{I_2}{V_2}\right|_{V_1=0} = \text{short-circuit output admittance}$$

For a given device, single transistor or cascode pair, these parameters may be specified as explicit functions of frequency, or, as is more often the case, as graphs of the real and imaginary parts, the conductance G and the susceptance B, versus frequency. The data sheet of the MC 1550 gives the y-parameters measured on the General Radio 1607A immittance bridge. Typical measured values are shown in Fig. 16-28. The internal feedback factor y_{12} is not shown because it was found to be less than 0.001 mA/V (m℧) and is neglected.

Let us consider the two-port network of Fig. 16-27b terminated at the output by a load admittance Y_L and driven by a current source I_s with source admittance Y_s. The equivalent admittance seen by the current source is $Y_{\text{eq}} = Y_s + Y_i$. In Prob. 16-36 we show that

$$Y_i = \frac{I_1}{V_1} = y_{11} - \frac{y_{12}y_{21}}{y_{22} + Y_L} \tag{16-54}$$

and the output admittance is

$$Y_o = \frac{I_2}{V_2} = y_{22} - \frac{y_{12}y_{21}}{y_{11} + Y_s} \tag{16-55}$$

Since $y_{12} \approx 0$, then $Y_i \approx y_{11}$ and $Y_o \approx y_{22}$.

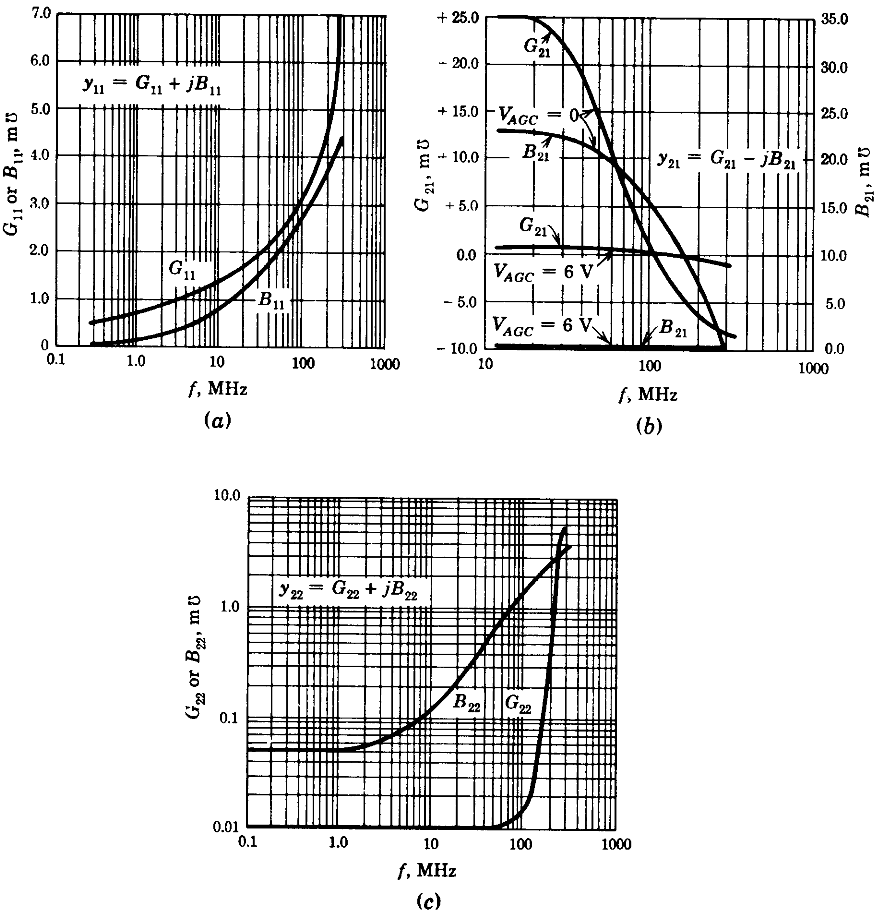


Fig. 16-28 The y-parameters of the MC 1550 for $V_{CC} = 6.0$ V and $V_{AGC} = 0$ V as functions of frequency. (a) The parameter y_{11}, (b) the parameter y_{21}, (c) the parameter y_{22}. (Courtesy of Motorola, Inc.)

The average power P_{av} delivered by the current source to the two-port is the power dissipated in the conductive part of Y_{eq}, or

$$P_{av} = \tfrac{1}{2}|V_1|^2 \operatorname{Re}\,[Y_{eq}]\dagger \tag{16-56}$$

If Re $[Y_{eq}]$ becomes negative at some frequency ω_1, the network absorbs negative power; in other words, power is supplied to the source by the network. We note from Eq. (16-54) that if $y_{12} \approx 0$ and Re $[y_{11}] > 0$ and Re $[Y_s] > 0$, the circuit cannot oscillate.

---

† Re $[Y_{eq}]$ means the real part of Y_{eq} and V_1 is the rms value of the input voltage.

The current gain A_I and voltage gain A_V are found in Prob. 16-37 to be given by

$$A_I = -\frac{I_2}{I_1} = -\frac{y_{21}Y_L}{y_{11}y_{22} - y_{12}y_{21} + y_{11}Y_L} \tag{16-57}$$

and

$$A_V = \frac{V_2}{V_1} = -\frac{y_{21}}{y_{22} + Y_L} \tag{16-58}$$

A Practical Tuned Amplifier A hybrid monolithic circuit which embodies the principles discussed above is indicated in Fig. 16-29. (For the moment, assume that the audio generator V_a is not present; $V_a = 0$.) The shaded block is the MC 1550 IC chip of Fig. 16-25. All other components are discrete elements added externally. Resistors R_1 and R_2 bias the diode $D1$ (and hence determine the collector current of $Q1$). These resistors also establish the bias voltage for $Q3$. Resistors R_3 and R_4 serve to "widen" the AGC voltage range from 120 mV to approximately 850 mV, thus rendering the AGC terminal less susceptible to external noise pickup.

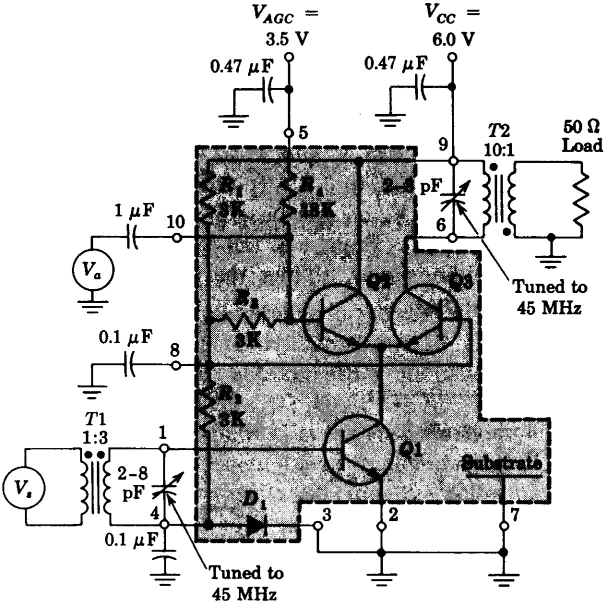

V_{AGC} =
3.5 V
V_{CC} =
6.0 V
0.47 μF
0.47 μF
5
9
T2
10:1
50 Ω
Load
R_1
3K
R_4
18K
2-8 pF
6
1 μF
10
V_a
R_3
3K
Q2
Q3
Tuned to
45 MHz
0.1 μF
8
R_2
3K
T1
1:3
1
V_s
2-8
pF
4
Q1
Substrate
D_1
3
2
7
0.1 μF
Tuned to
45 MHz

Fig. 16-29 A practical 45-MHz tuned amplifier (with $V_a = 0$), or an RF modulator if $V_a \neq 0$. (Courtesy of Motorola Semiconductor, Inc.)

The source V_s is a 45-MHz RF (radio-frequency) generator whose resistance is 50 Ω. The transformers are wound with No. 32 wire on T12-2 cores; T1 with 6:18 turns has a magnetizing inductance $L_M = 1.1\ \mu\text{H}$, and T2 with 30:3 turns gives $L_M = 2.5\ \mu\text{H}$. The variable capacitors in Fig. 16-29 are adjusted so as to resonate with these inductors at 45 MHz.

For maximum power gain through an amplifier the source admittance and load admittance must be selected to be the complex conjugates of the input admittance $Y_i \approx y_{11}$ and of the output admittance $Y_o \approx y_{22}$, respectively. In place of transformers, LC networks may be used to obtain this matching. In Fig. 16-30, the input network consisting of C_1, C_2, and L_1 transforms the 50-Ω resistance of the source into the complex conjugate of y_{11}. The values of C_1, C_2, and L_1 are calculated using techniques in Ref. 8. The output network consists of C_3, C_4, and L_2 and transforms the 50-Ω load resistance into the complex conjugate of y_{22}. The center frequency for this amplifier is 60 MHz, the bandwidth is 500 kHz, and the power gain is measured to be 30 dB.

Amplitude Modulator The RF carrier signal V_s may be varied in amplitude by changing the AGC voltage. Hence, if an audio signal V_a is applied to terminal 10 in Fig. 16-29 so as to modify the AGC voltage, the output will be the amplitude-modulated waveform indicated in Fig. 4-27.

The gain of the $Q1$-$Q3$ cascode is proportional to $|y_{21}|$, as indicated in Eq. (16-58). The parameter y_{21} depends upon the collector current of $Q3$, which can be varied by changing V_{AGC}. Figure 16-31 shows the variation of $|y_{21}|$ versus V_{AGC} at the frequency of 45 MHz. From the curve we see that between $V_{AGC} = 2.75$ V and $V_{AGC} = 4.25$ V, $|y_{21}|$ is linear with V_{AGC}. By biasing the AGC line to 3.5 V (point B on the curve) and impressing an audio sinusoidal signal V_a on the base of $Q2$ (as indicated in Fig. 16-29), $|y_{21}|$ will vary sinusoidally. From Eq. (16-58) the gain A_V will also vary sinusoidally with the audio signal. Thus the amplifier output will be an amplitude-modulated signal.

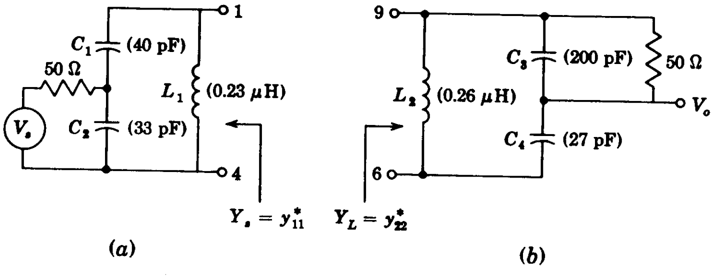


Fig. 16-30 Matching networks for maximum power transfer in a 60-MHz tuned amplifier (a) at the input and (b) at the output.

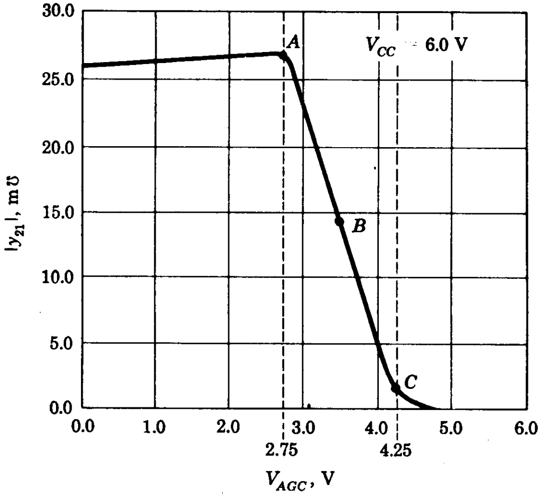


Fig. 16-31 Variation of $|y_{21}|$ versus V_{AGC} at 45 MHz for the MC 1550. (Courtesy of Motorola Semiconductor, Inc.)

16-10 A CASCODE VIDEO AMPLIFIER

A video amplifier, as opposed to a tuned amplifier, must amplify signals over a wide band of frequencies, say up to 20 MHz. The MC 1550 can be used as a cascode video amplifier (Fig. 16-32*a*) by avoiding tuned input and output circuits. Between pins 1 and 4, 50 Ω is inserted in order to properly terminate the coaxial cable carrying the video signal. This small resistance has negligible effect on the biasing of $Q1$. The load R_L is placed directly in the collector lead of $Q3$.

The small-signal analysis of this video amplifier can be made using the approximate equivalent circuit of Fig. 16-32*c*. If both transistors $Q2$ and $Q3$ are operating in their active region, the collector of $Q1$ sees the very small input resistance $(r_{e2} \| r_{e3})$ of two common-base stages in parallel. We can represent $Q1$ with its hybrid-Π model, and due to the very low load on $Q1$, we can neglect the effect of C_c. The video output is taken from the collector of $Q3$, which is operating as a common-base stage. We shall assume that $Q3$ can be represented by a current source $\alpha_3 I_3$, where I_3 is the emitter signal current of $Q3$ and $\alpha_3 \approx 1$ and is independent of frequency over the band of frequencies under consideration. C_s represents the capacitance from the collector of $Q1$ and $Q3$ to the substrate (ground).

From Fig. 16-32*c* we find (Prob. 16-42)

$$\frac{V_o}{V_i} = -\frac{\alpha_3 g_m}{r_{e3} r_{bb'}} \frac{1}{\left(\dfrac{1}{r_{bb'}} + \dfrac{1}{r_{b'e}} + sC_e\right)\left(\dfrac{1}{r_{e2}} + \dfrac{1}{r_{e3}} + sC_s\right)\left[\dfrac{1}{R_L} + s(C_s + C_L)\right]} \tag{16-59}$$

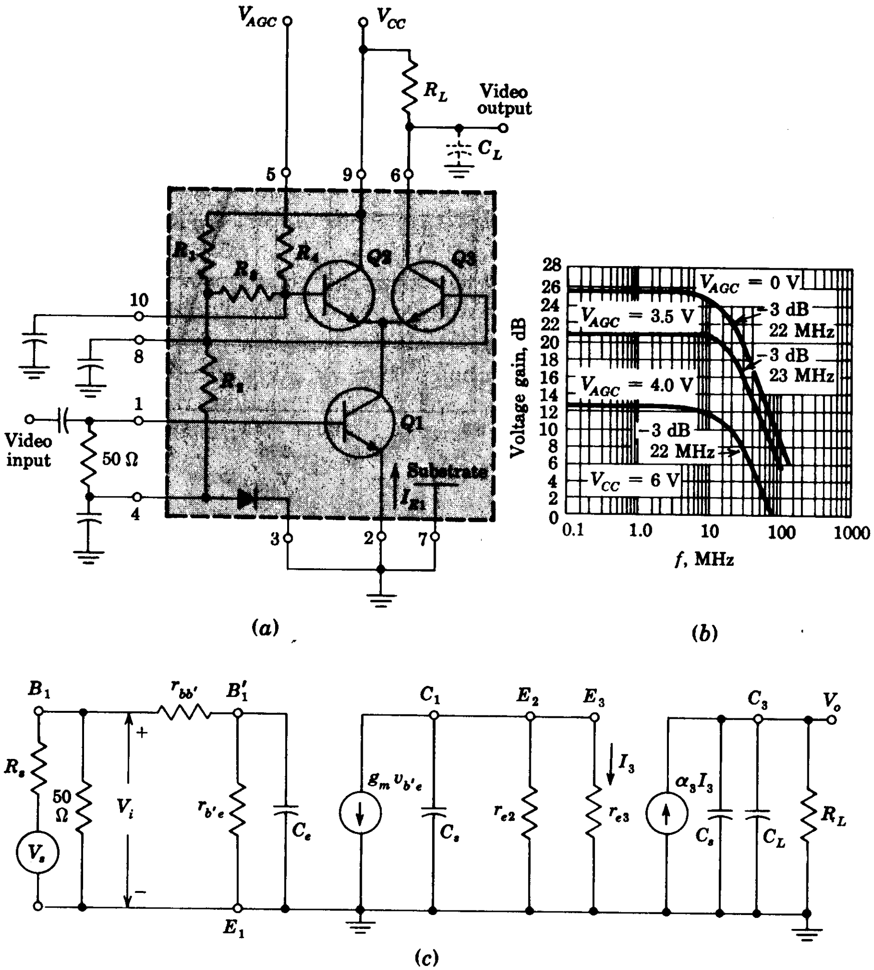

V_{AGC}
V_{CC}
R_L
Video output
C_L
5
9
6
10
8
1
4
3
2
7
R_1
R_2
R_3
R_4
Q1
Q2
Q3
Substrate
I_{E1}
Video input
50 Ω
(a)
Voltage gain, dB
V_{AGC} = 0 V
V_{AGC} = 3.5 V
V_{AGC} = 4.0 V
V_{CC} = 6 V
−3 dB 22 MHz
−3 dB 23 MHz
−3 dB 22 MHz
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28
0.1 1.0 10 100 1000
f, MHz
(b)
B_1
$r_{bb'}$
B'_1
R_s
50 Ω
V_s
V_i
+
−
$r_{b'e}$
C_e
E_1
C_1
$g_m v_{b'e}$
C_s
E_2
r_{e2}
E_3
I_3
r_{e3}
$\alpha_3 I_3$
C_s
C_3
C_L
R_L
V_o
(c)

Fig. 16-32 (*a*) The MC 1550 used as a video amplifier; (*b*) frequency response for three different values of V_{AGC}; (*c*) approximate small-signal equivalent circuit.

From Eq. (3-14) the emitter-base-diode incremental resistance r_e (r_{e2} or r_{e3}) is given by $r_e = \eta V_T / I_E$, where I_E is the quiescent emitter current.

EXAMPLE Design a video amplifier, using the circuit of Fig. 16-32 and the MC 1550, to provide voltage gain $A_V = V_o/V_i = -25$ and bandwidth greater than 20 MHz when $V_{AGC} = 0$. Assume $V_{CC} = 6$ V, $h_{fe} = 50$, $r_{bb'} = 50\ \Omega$, $C_s = 5$ pF, $C_L = 5$ pF, $I_{E1} = 1$ mA, and $f_T = 900$ MHz.

Solution The low-frequency voltage gain V_o/V_i is found from Eq. (16-59) by letting $s = 0$. When $V_{AGC} = 0$, transistor $Q2$ is cut off and all the collector current of $Q1$ flows through $Q3$. Since $I_{E2} = 0$, then $r_{e2} = \eta V_T/I_{E2} = \infty$ and

$$r_{e3} = \eta \frac{V_T}{I_{E3}} = \frac{52 \text{ mV}}{1 \text{ mA}} = 52\ \Omega \text{ at } 25°\text{C}$$

From Sec. 11-2 we obtain

$$g_m = \frac{I_{E1}}{V_T} = \frac{1}{26} = 38.5 \times 10^{-3} \text{ A/V}$$

$$r_{b'e} = \frac{h_{fe}}{g_m} = \frac{50}{38.5 \times 10^{-3}}\ \Omega = 1.30 \text{ K}$$

and

$$C_e = \frac{g_m}{2\pi f_T} = \frac{38.5 \times 10^{-3}}{2 \times 3.14 \times 900 \times 10^{-6}} \text{ F} = 6.80 \text{ pF}$$

If we assume $\alpha_3 \approx 1$, we find from Eq. (16-59)

$$A_{V_o} = \left.\frac{V_o}{V_i}\right|_{s=0} = -\frac{38.5 \times 10^{-3}}{52 \times 50} \frac{1}{\left(\frac{1}{50} + \frac{1}{1{,}300}\right) \times \frac{1}{52} \times \frac{1}{R_L}}$$

$$= -37.2 \times 10^{-3} R_L$$

Thus

$$A_{V_o} = -25 = -37.2 \times 10^{-3} \times R_L$$

or

$$R_L = \frac{25}{37.2} \times 10^3 = 675\ \Omega$$

The voltage transfer function of Eq. (16-59) has three poles, and the corresponding 3-dB frequencies are

$$f_1 = \frac{1}{2\pi R_L(C_s + C_L)} = \frac{1}{2 \times 3.14 \times 675 \times 10 \times 10^{-12}} \text{ Hz} = 23.6 \text{ MHz}$$

$$f_2 = \frac{1}{2\pi C_e(r_{b'e} \| r_{bb'})} = \frac{1}{2 \times 3.14 \times 6.80 \times 10^{-12} \times 48} \text{ Hz} = 490 \text{ MHz}$$

$$f_3 = \frac{1}{2\pi C_s(r_{e2} \| r_{e3})} = \frac{1}{2 \times 3.14 \times 5 \times 10^{-12} \times 52} \text{ Hz} = 610 \text{ MHz}$$

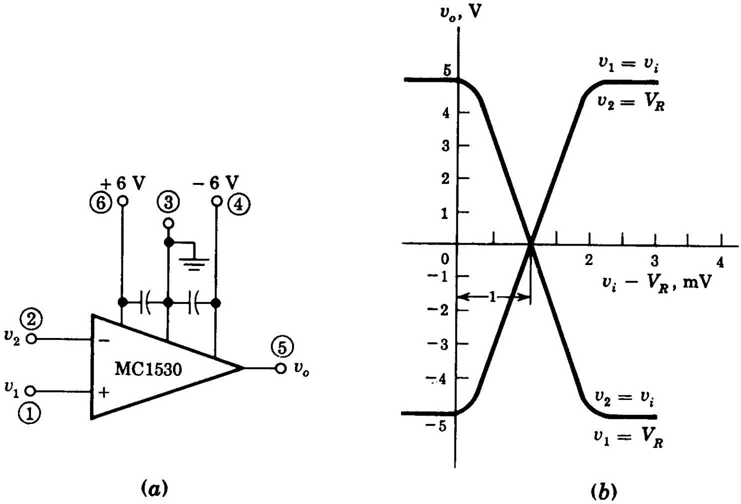
We conclude that f_1 is a dominant pole and $f_H \approx f_1 = 23.6$ MHz. Figure 16-32*b* shows measured data for three values of V_{AGC}. We see that, although we used a simplified model to analyze the circuit, we obtained excellent agreement with experiment.

II. NONLINEAR ANALOG SYSTEMS

16-11 COMPARATORS

With the exception of amplitude modulation and automatic gain control, all the systems discussed thus far in this chapter have operated linearly. The remainder of the chapter is concerned with nonlinear OP AMP functions.

The comparator, introduced in Sec. 4-6, is a circuit which compares an input signal $v_i(t)$ with a reference voltage V_R. When the input v_i exceeds V_R, the comparator output v_o takes on a value which is very different from the magnitude of v_o when v_i is smaller than V_R. The DIFF AMP input-output curve of Fig. 15-9 approximates this comparator characteristic. Note that the total input swing between the two extreme output voltages is $\sim 8V_T = 200$ mV. This range may be reduced considerably by cascading two DIFF AMPS as in Fig. 15-11. This MC 1530 OP AMP serves as a comparator if connected open-loop, as shown in Fig. 16-33*a*. The transfer characteristic is given in Fig. 16-33*b*, and it is now observed that the change in output state takes place with a variation in input of only 2 mV. Note that the input offset voltage contributes an error in the point of comparison between v_i and V_R of the order of 1 mV. The reference V_R may be any voltage, provided that it does not exceed the maximum common-mode range.



(*a*) (*b*)

Fig. 16-33 (*a*) The MC 1530 operational amplifier as a comparator. (*b*) The transfer characteristic.

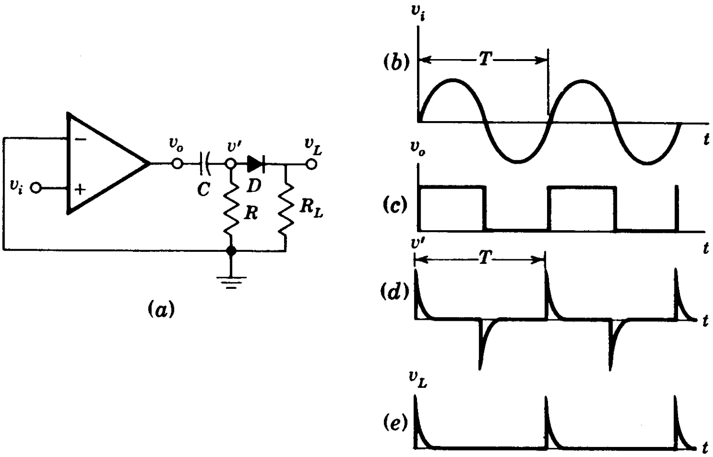


Fig. 16-34 A zero-crossing detector converts a sinusoid v_i into a square wave v_o. The pulse waveforms v' and v_L result if v_o is fed into a short time-constant *RC* circuit in cascade with a diode clipper.

Zero-crossing Detector If V_R is set equal to zero, the output will respond almost discontinuously every time the input passes through zero. Such an arrangement is called a *zero-crossing detector.*

Some of the most important systems using comparators will now be listed. Other applications are discussed in Secs. 16-5 and 17-19.

Square Waves from a Sine Wave If the input to an OP AMP comparator is a sine wave, the output is a square wave. If a zero-crossing detector is used, a symmetrical square wave results, as shown in Fig. 16-34*c*. This idealized waveform has vertical sides which, in reality, should extend over a range of a few millivolts of input voltage v_i.

Timing-markers Generator from a Sine Wave The square-wave output v_o of the preceding application is applied to the input of an *RC* series circuit. If the time constant *RC* is very small compared with the period *T* of the sine-wave input, the voltage v' across *R* is a series of positive and negative pulses, as indicated in Fig. 16-34*d*. If v' is applied to a clipper with an ideal diode (Fig. 16-34*a*), the load voltage v_L contains only positive pulses (Fig. 16-34*e*). Thus the sinusoid has been converted into a train of positive pulses whose spacing is *T*. These may be used for timing markers (on the sweep voltage of a cathode-ray tube, for example).

Phasemeter The phase angle between two voltages can be measured by a method based on the circuit of Fig. 16-34. Both voltages are converted into pulses, and the time interval between the pulse of one wave and that obtained from the second sine wave is measured. This time interval is proportional to the phase difference. Such a phasemeter can measure angles from 0 to 360°.

Amplitude-distribution Analyzer A comparator is a basic building block in a system used to analyze the amplitude distribution of the noise generated in an active device or the voltage spectrum of the pulses developed by a nuclear-radiation detector, etc. To be more specific, suppose that the output of the comparator is 10 V if $v_i > V_R$ and 0 V if $v_i < V_R$. Let the input to the comparator be noise. A dc meter is used to measure the average value of the output square wave. For example, if V_R is set at zero, the meter will read 10 V, which is interpreted to mean that the probability that the amplitude is greater than zero is 100 percent. If V_R is set at some value V'_R and the meter reads 7 V, this is interpreted to mean that the probability that the amplitude of the noise is greater than V'_R is 70 percent, etc. In this way the cumulative amplitude probability distribution of the noise is obtained by recording meter readings as a function of V_R.

Pulse-time Modulation If a periodic sweep waveform is applied to a comparator whose reference voltage V_R is not constant but rather is modulated by an audio signal, it is possible to obtain a succession of pulses whose relative spacing reflects the input information. The result is a *time-modulation system* of communication.

16-12 SAMPLE-AND-HOLD CIRCUITS[9]

A typical data-acquisition system receives signals from a number of different sources and transmits these signals in suitable form to a computer or a communication channel. A multiplexer (Sec. 17-5) selects each signal in sequence, and then the analog information is converted into a constant voltage over the gating-time interval by means of a *sample-and-hold circuit*. The constant output of the sample-and-hold may then be converted to a digital signal by means of an analog-to-digital (A/D) converter (Sec. 17-20) for digital transmission.

A sample-and-hold circuit in its simplest form is a switch S in series with a capacitor, as in Fig. 16-35*a*. The voltage across the capacitor tracks the input signal during the time T_g when a logic control gate closes S, and holds the instantaneous value attained at the end of the interval T_g when the control gate opens S. The switch may be a relay (for very slow waveforms), a sampling diode-bridge gate (Sec. 4-7), a bipolar transistor switch,[10] or a MOSFET controlled by a gating signal. The MOSFET makes an excellent

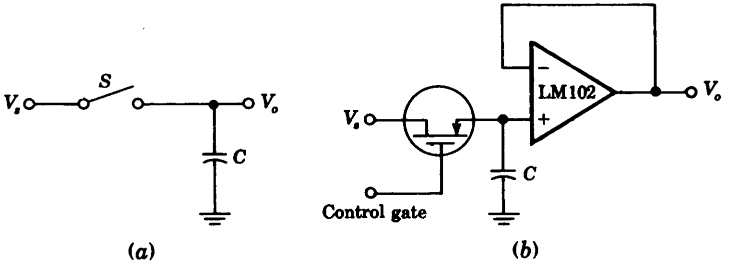


Fig. 16-35 Sample-and-hold circuit. (a) Schematic, (b) practical.

chopper because its *offset voltage* when ON (~5 μV) is much smaller than that of a bipolar junction transistor.

The circuit shown in Fig. 16-35*b* is one of the simplest practical sample-and-hold circuits. A negative pulse at the gate of the *p*-channel MOSFET will turn the switch ON, and the holding capacitor C will charge with a time constant $R_{ON}C$ to the instantaneous value of the input voltage. In the absence of a negative pulse, the switch is turned OFF and the capacitor is isolated from any load through the LM 102 OP AMP. Thus it will hold the voltage impressed upon it. It is recommended that a capacitor with polycarbonate, polyethylene, or Teflon dielectric be used. Most other capacitors do not retain the stored voltage, due to a polarization phenomenon[11] which causes the stored voltage to decrease with a time constant of several seconds. Even if the polarization phenomenon does not occur, the OFF current of the switch (~1 nA) and the bias current of the OP AMP will flow through C. Since the maximum input bias current for the LM 102 is 10 nA, it follows that with a 10-μF capacitance the drift rate during the HOLD period will be less than 1 mV/s.

Two additional factors influence the operation of the circuit: the reaction time, called *aperture time* (typically less than 100 ns), which is the delay between the time that the pulse is applied to the switch and the actual time the switch closes, and the *acquisition time*, which is the time it takes for the capacitor to change from one level of holding voltage to the new value of input voltage after the switch has closed.

When the hold capacitor is larger than 0.05 μF, an isolation resistor of approximately 10 K should be included between the capacitor and the + input of the OP AMP. This resistor is required to protect the amplifier in case the output is short-circuited or the power supplies are abruptly shut down while the capacitor is charged.

16-13 PRECISION AC/DC CONVERTERS[12]

If a sinusoid whose peak value is less than the threshold or cutin voltage V_γ (~0.6 V) is applied to the rectifier circuit of Fig. 4-6, we see that the output

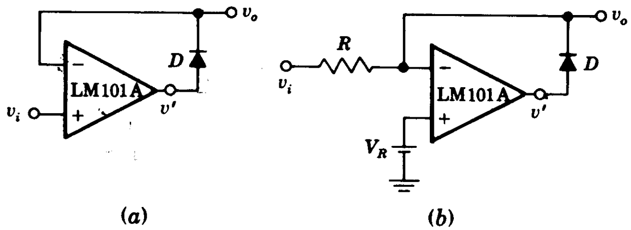


Fig. 16-36 (a) A precision diode. (b) A precision clamp.

is zero for all times. In order to be able to rectify millivolt signals, it is clearly necessary to reduce V_γ. By placing the diode in the feedback loop of an OP AMP, the cutin voltage is divided by the open-loop gain A_V of the amplifier. Hence V_γ is virtually eliminated and the diode approaches the ideal rectifying component. If in Fig. 16-36*a* the input v_i goes positive by at least V_γ/A_V, then v' exceeds V_γ and D conducts. Because of the virtual connection between the noninverting and inverting inputs (due to the feedback with D ON), $v_o \approx v_i$. Therefore the circuit acts as a voltage follower for positive signals (in excess of approximately 0.1 mV). When v_i swings negatively, D is OFF and no current is delivered to the external load except for the small bias current of the LM 101A.

Precision Clamp By modifying the circuit of Fig. 16-36*a*, as indicated in Fig. 16-36*b*, an almost ideal clamp (Sec. 4-5) is obtained. If $v_i < V_R$, then v' is positive and D conducts. As explained above, under these conditions the output equals the voltage at the noninverting terminal, or $v_o = V_R$. If $v_i > V_R$, then v' is negative, D is OFF, and $v_o = v_i$. In summary: The output follows the input for $v_i > V_R$ and v_o is clamped to V_R if v_i is less than V_R by about 0.1 mV. When D is reverse-biased in Fig. 16-36*a* or *b*, a large differential voltage may appear between the inputs and the OP AMP must be able to withstand this voltage. Also note that when $v_i > V_R$, the input stage saturates because the feedback through D is missing.

Fast Half-wave Rectifier By adding R' and $D2$ to Fig. 16-36*b* and setting $V_R = 0$, we obtain the circuit of Fig. 16-37*a*. If v_i goes negative, D1 is ON, $D2$ is OFF, and the circuit behaves as an inverting OP AMP, so that $v_o = -(R'/R)v_i$. If v_i is positive, $D1$ is OFF and $D2$ is ON. Because of the feedback through $D2$, a virtual ground exists at the input and $v_o = 0$. If v_i is a sinusoid, the circuit performs half-wave rectification. Because the amplifier does not saturate, it can provide rectification at frequencies up to 100 kHz.

An equivalent alternative configuration to that in Fig. 16-37*a* is to ground the left-hand side of R and to impress v_i at the noninverting terminal. The half-wave-rectified output now has a peak value of $(R + R')/R$ times the

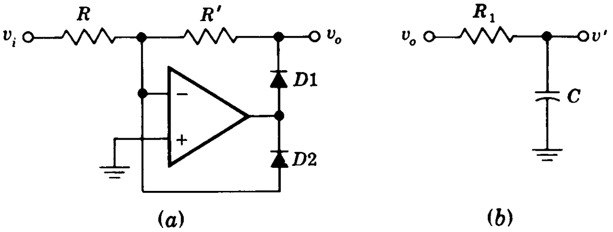


Fig. 16-37 (a) A half-wave rectifier. (b) A low-pass filter which can be cascaded with the circuit in (a) to obtain an average detector.

maximum sinusoidal input voltage. A full-wave system is indicated in Prob. 16-43.

Active Average Detector Consider the circuit of Fig. 16-37a to be cascaded with the low-pass filter of Fig. 16-37b. If v_i is an amplitude-modulated carrier (Fig. 4-27), the R_1C filter removes the carrier and v' is proportional to the average value of the audio signal. In other words, this configuration represents an *average detector.*

Active Peak Detector If a capacitor is added at the output of the precision diode of Fig. 16-36a, a peak detector results. The capacitor in Fig. 16-38a will hold the output at $t = t'$ to the most positive value attained by the input v_i prior to t', as indicated in Fig. 16-38b. This operation follows from the fact that if $v_i > v_o$, the OP AMP output v' is positive, so that D conducts. The capacitor is then charged through D (by the output current of the amplifier) to the value of the input because the circuit is a voltage follower. When v_i falls below the capacitor voltage, the OP AMP output goes negative

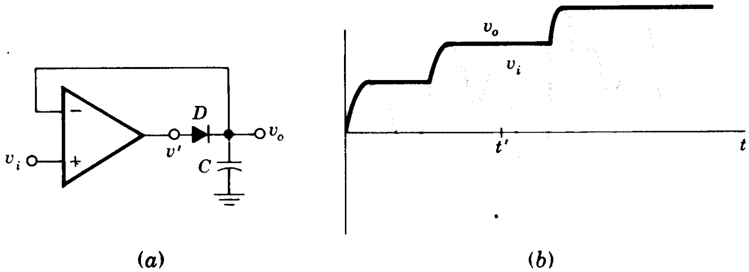


Fig. 16-38 (a) A positive peak detector. (b) An arbitrary input waveform v_i and the corresponding output v_o.

and the diode becomes reverse-biased. Thus the capacitor gets charged to the most positive value of the input.

This circuit is a special case of a sample-and-hold circuit, and the capacitor leakage current considerations given in Sec. 16-12 also apply to this configuration. If the output is loaded, a buffer voltage follower should be used to prevent the load from discharging C. To reset the circuit, a low-leakage switch such as a MOSFET gate must be placed across the capacitor.

16-14 LOGARITHMIC AMPLIFIERS[12]

In Fig. 16-39*a* there is indicated an OP AMP with the feedback resistor R' replaced by the diode $D1$. This amplifier is used when it is desired to have the output voltage proportional to the logarithm of the input voltage.

From Eq. (3-9) the volt-ampere diode characteristic is

$$I_f = I_o(\epsilon^{V_f/\eta V_T} - 1) \approx I_o\epsilon^{V_f/\eta V_T}$$

provided that $V_f/\eta V_T \gg 1$ or $I_f \gg I_o$. Hence

$$V_f = \eta V_T(\ln I_f - \ln I_o) \tag{16-60}$$

Since $I_f = I_s = V_s/R$ due to the virtual ground at the amplifier input, then

$$V_o = -V_f = -\eta V_T\left(\ln \frac{V_s}{R} - \ln I_o\right) \tag{16-61}$$

We note from Eq. (16-61) that the output voltage V_o is temperature-dependent due to the scale factor ηV_T and to the saturation current I_o. Both temperature effects can be reduced by using the circuit of Fig. 16-39*b*, where the diodes $D1$ and $D2$ are matched, R_T is temperature-dependent, and the constant source I is independent of T.

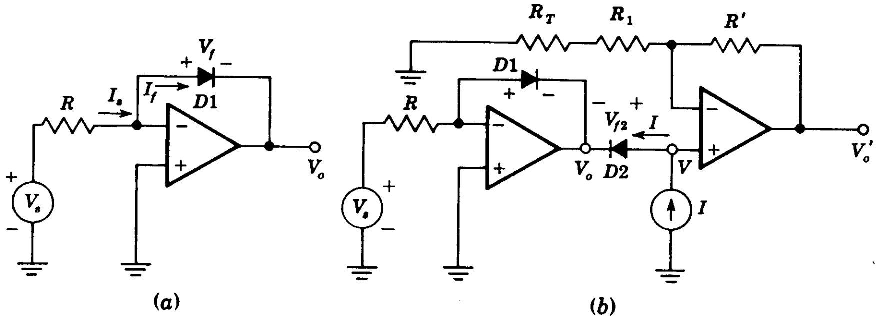


Fig. 16-39 (*a*) Logarithmic amplifier for positive input voltage V_s. (*b*) Temperature-compensated amplifier.

We have for this circuit, and using Eq. (16-61),

$$V = V_{f2} + V_o = \eta V_T \left(\ln I - \ln I_o - \ln \frac{V_s}{R} + \ln I_o \right) = -\eta V_T \ln \frac{V_s}{RI}$$

Thus the output voltage V_o' becomes

$$V_o' = -\frac{R_T + R_1 + R'}{R_1 + R_T} \eta V_T \ln \frac{V_s}{RI} \tag{16-62}$$

The temperature dependence of R_T is selected to compensate approximately for the factor ηV_T in Eq. (16-62).

Logarithmic Amplifier Using Matched Transistors Instead of two matched diodes, it is possible to use a matched pair of transistors connected as in Fig. 16-40*a* to remove from the expression for V_o' the factor η, whose value normally depends on the current flowing through the diode. In Fig. 16-40*a*, *Q*1 is used as the feedback element around the first OP AMP. If we neglect $V_{BE1} - V_{BE2}$ with respect to V_{CC}, and since $I_{B2} \ll I_{C2}$, then

$$I_{C2} \approx \frac{V_{CC}}{R_6} \qquad \text{and} \qquad I_{C1} = \frac{V_s}{R_1 + R_4} = \frac{V_s}{2R_1} \tag{16-63}$$

From Eq. (15-21) it follows that

$$V_{BE1} - V_{BE2} = V_T \ln I_{C1} - V_T \ln I_{C2} = V_T \ln \left(\frac{V_s}{2R_1} \frac{R_6}{V_{CC}} \right) \tag{16-64}$$

Since the base of *Q*1 is grounded, the negative of the above voltage appears at the noninverting terminal of the second operational amplifier, whose gain is determined by resistors R_7 and R_8. Hence

$$V_o = -V_T \frac{R_7 + R_8}{R_7} \ln \left(\frac{V_s}{2R_1} \frac{R_6}{V_{CC}} \right) \tag{16-65}$$

The above transfer function of the amplifier is plotted in Fig. 16-40*b* for various operating temperatures. It is seen that the dynamic range extends over 5 mV to 50 V of input voltage, or 80 dB. From Eq. (16-65),

$$\frac{dV_o}{d(\ln V_s)} = -V_T \left(\frac{R_7 + R_8}{R_7} \right) = -0.026 \times \frac{43.8}{0.52} = -2.20$$

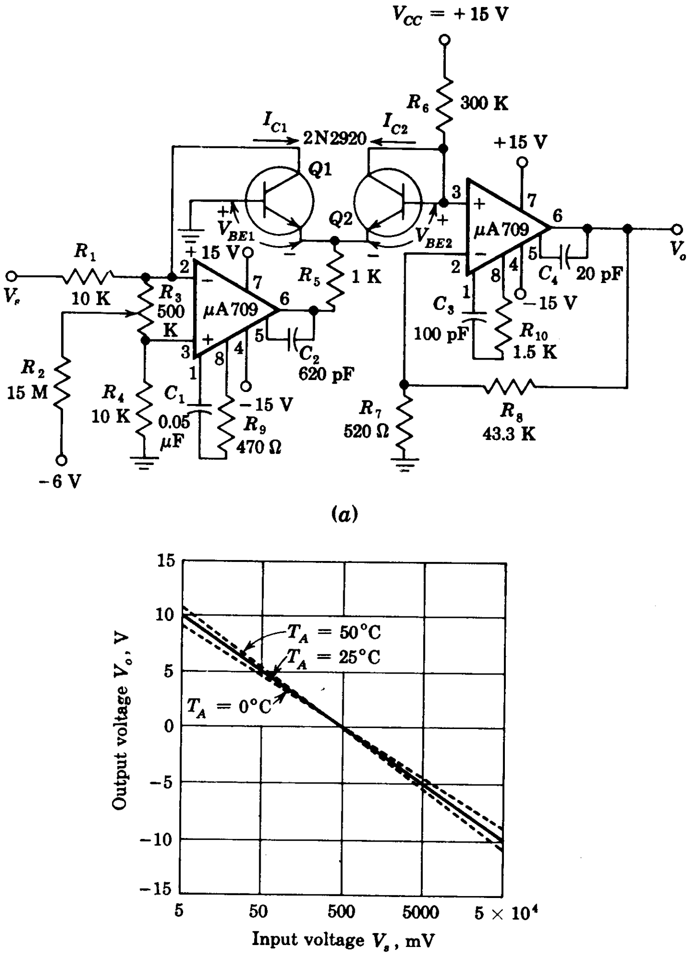
which is in excellent agreement with the slope obtained from Fig. 16-40*b*.

Antilog Amplifier The amplifiers discussed above give an output V_o proportional to the natural logarithm of the input V_s, or

$$V_o = K_1 \ln K_2 V_s \tag{16-66}$$

Sometimes we desire an output proportional to the antilogarithm ($\ln^{-1}$) of the input; that is,

$$V_o = K_3 \ln^{-1} K_4 V_s \tag{16-67}$$


V_{CC} = +15 V
R_6 300 K
I_{C1}
I_{C2}
2N2920
Q1
Q2
+15 V
3
+
μA709
7
6
V_o
V_{BE1}
V_{BE2}
R_1
10 K
V_s
+15 V
2
R_3 500 K
μA709
R_5 1 K
5
C_4 20 pF
−15 V
8
4
C_3 100 pF
R_{10} 1.5 K
C_2 620 pF
R_2 15 M
R_4 10 K
C_1 0.05 μF
−15 V
R_9 470 Ω
R_7 520 Ω
R_8 43.3 K
−6 V
15
10
5
0
−5
−10
−15
Output voltage V_o, V
T_A = 50°C
T_A = 25°C
T_A = 0°C
5
50
500
5000
5 × 10⁴
Input voltage V_s, mV

(a)

(b)

Fig. 16-40 (a) Logarithmic amplifier. (b) Transfer characteristic. (Courtesy of Fairchild Semiconductor, Inc.)

The circuit shown in Fig. 16-41 can be used as an antilog amplifier. If we assume infinite input resistance for A_1 and A_2 as well as zero differential input voltage for each operational amplifier, we obtain

$$V_2 = -V_f + V_1 = -\eta V_T(\ln I_f - \ln I_o) + \frac{R_1}{R_1 + R_2} V_s \qquad (16\text{-}68)$$

and since V_2 is the negative of the voltage across $D2$,

$$V_2 = -\eta V_T(\ln I_2 - \ln I_o) \qquad (16\text{-}69)$$

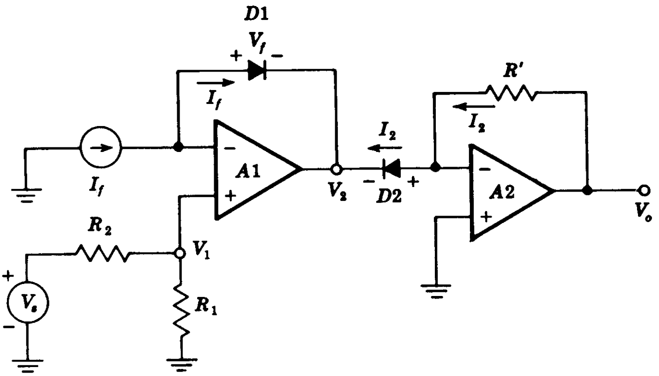


Fig. 16-41 Antilog amplifier.

Combining Eqs. (16-68) and (16-69) yields

$$V_s \frac{R_1}{R_1 + R_2} = \eta V_T \ln \frac{I_f}{I_2} = \eta V_T \ln \frac{I_f R'}{V_o} \tag{16-70}$$

because $V_o = I_2 R'$. Finally, from Eq. (16-70) it follows that

$$V_o = R' I_f \ln^{-1}\left[-V_s\left(\frac{R_1}{R_1 + R_2}\frac{1}{\eta V_T}\right)\right] \tag{16-71}$$

Equation (16-71) is of the form given in Eq. (16-67).

We show in Prob. 16-45 that it is possible to raise the input V_s to an arbitrary power by combining log and antilog amplifiers.

Logarithmic Multiplier The log and antilog amplifiers can be used for the multiplication or division of two analog signals V_{s1} and V_{s2}. In Fig. 16-42

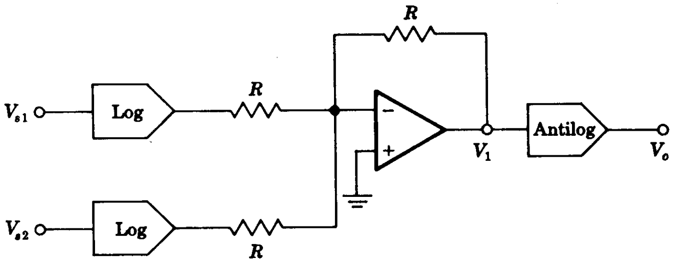


Fig. 16-42 Logarithmic multiplier of two analog signals ($V_o = KV_{s1}V_{s2}$).

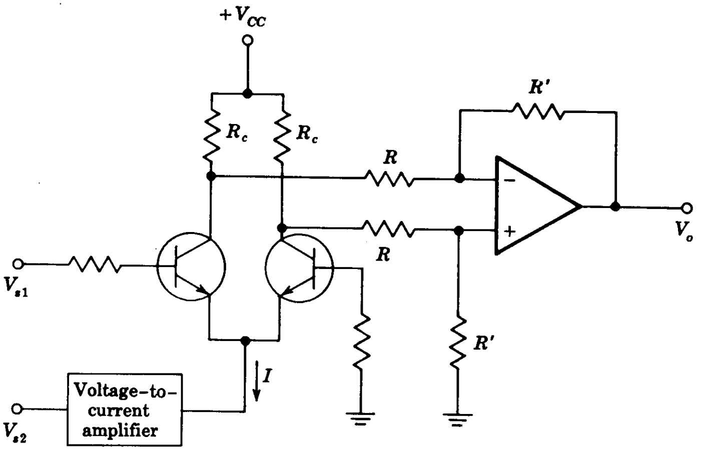


Fig. 16-43 Variable transconductance multiplier ($V_o = KV_{s1}V_{s2}$).

the logarithm of each input is taken, then the two logarithms are added, and finally the antilog of the sum yields the product of the two inputs. Thus

$$V_1 = K_1 \ln V_{s1} + K_1 \ln V_{s2} = K_1 \ln V_{s1}V_{s2} \tag{16-72}$$

and

$$V_o = K_2 \ln^{-1} K_3V_1 = K_2 \ln^{-1} (K_3K_1 \ln V_{s1}V_{s2}) \tag{16-73}$$

If $K_3K_1 = 1$, then

$$V_o = K_2V_{s1}V_{s2} \tag{16-74}$$

The input signals can be divided if we subtract the logarithm of V_{s1} from that of V_{s2} and then take the antilog. We must point out that the logarithmic multiplier or divider is useful for unipolar inputs only. This is often called one-quadrant operation. Other techniques[13] are available for the accurate multiplication of two signals.

Differential Amplifier Multiplier From Eqs. (15-24) and (15-23) we observe that the output voltage of a differential amplifier depends on the current source I. If V_{s1} is applied to one input and V_{s2} is used to vary I, as in Fig. 16-43, the output will be proportional to the product of the two signals $V_{s1}V_{s2}$. The device AD 530 manufactured by Analog Devices, Inc., is a completely monolithic multiplier/divider with basic accuracy of 1 percent and bandwidth of 1 MHz. As a multiplier, the AD 530 has the transfer function $XY/10$ and as a divider $+10Z/X$. The X, Y, and Z input levels are ±10 V for multiplication and the output is ±10 V at 5 mA. As a divider, operation is restricted to two quadrants (where X is negative) only.

16-15 WAVEFORM GENERATORS[9]

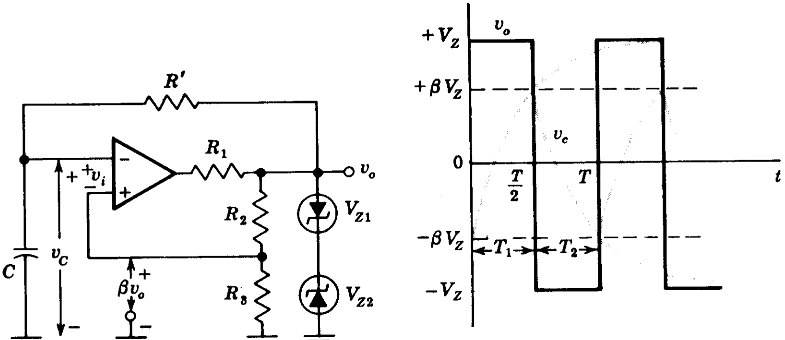
The operational amplifier comparator, together with an integrator, can be used to generate a square wave, a pulse, or a triangle waveform, as we now demonstrate.

Square-wave Generator In Fig. 16-44*a*, the output v_o is shunted to ground by two Zener diodes connected back to back and is limited to either $+V_{Z2}$ or $-V_{Z1}$, if $V_\gamma \ll V_Z$ (Fig. 4-11). A fraction $\beta = R_3/(R_2 + R_3)$ of the output is fed back to the noninverting input terminal. The differential input voltage v_i is given by

$$v_i = v_c - \beta v_o \tag{16-75}$$

From the transfer characteristic of the comparator given in Fig. 16-33 we see that if v_i is positive (by at least 1 mV), then $v_o = -V_{Z1}$, whereas if v_i is negative (by at least 1 mV), then $v_o = +V_{Z2}$. Consider an instant of time when $v_i < 0$ or $v_c < \beta v_o = \beta V_{Z2}$. The capacitor C now charges exponentially toward V_{Z2} through the integrating $R'C$ combination. The output remains constant at V_{Z2} until v_c equals $+\beta V_{Z2}$, at which time the comparator output reverses to $-V_{Z1}$. Now v_c charges exponentially toward $-V_{Z1}$. The output voltage v_o and capacitor voltage v_c waveforms are shown in Fig. 16-44*b* for the special case $V_{Z1} = V_{Z2} = V_Z$. If we let $t = 0$ when $v_c = -\beta V_Z$ for the first half cycle, we have

$$v_c(t) = V_Z[1 - (1 + \beta)\epsilon^{-t/R'C}] \tag{16-76}$$



(*a*) (*b*)

Fig. 16-44 (a) A square-wave generator. (b) Output and capacitor voltage waveforms.

Since at $t = T/2$, $v_c(t) = +\beta V_Z$, we find T, solving Eq. (16-76), to be given by

$$T = 2R'C \ln \frac{1 + \beta}{1 - \beta} \tag{16-77}$$

Note that T is independent of V_Z.

This square-wave generator is particularly useful in the frequency range 10 Hz to 10 kHz. At higher frequencies the delay time of the operational amplifier as it moves out of saturation, through its linear range, and back to saturation in the opposite direction, becomes significant. Also, the slew rate of the operational amplifier limits the slope of the output square wave. The frequency stability depends mainly upon the Zener-diode stability and the capacitor, whereas waveform symmetry depends on the matching of the two Zener diodes. If an unsymmetrical square wave is desired, then $V_{Z1} \neq V_{Z2}$.

The circuit will operate in essentially the same manner as described above if $R_1 = 0$ and the avalanche diodes are omitted. However, now the amplitude of the square wave depends upon the power supply voltage (±5.8 V for the MC 1530, using ±6 V supplies as in Sec. 15-5).

The circuit of Fig. 16-44 is called an *astable multivibrator* because it has two quasistable states. The output remains in one of these states for a time T_1 and then abruptly changes to the second state for a time T_2, and the cycle of period $T = T_1 + T_2$ repeats.

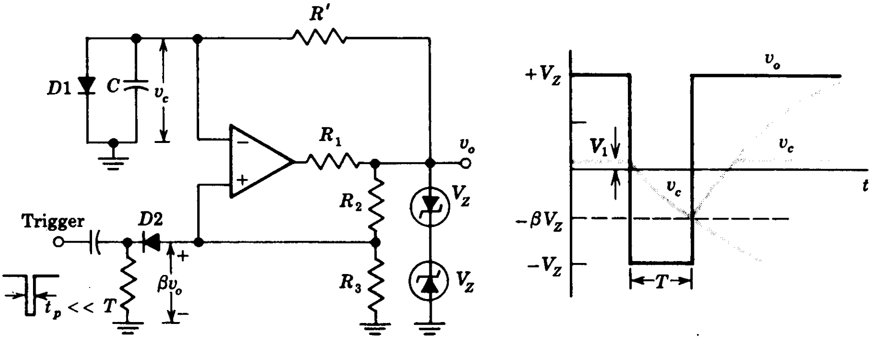
Pulse Generator A *monostable multivibrator* has one stable state and one quasistable state. The circuit remains in its stable state until a triggering signal causes a transition to the quasistable state. Then, after a time T, the circuit returns to its stable state. Hence a single pulse has been generated, and the circuit is referred to as a *one-shot.*

The square-wave generator of Fig. 16-44 is modified in Fig. 16-45 to operate as a monostable multivibrator by adding a diode ($D1$) clamp across C and by introducing a narrow negative triggering pulse through $D2$ to the noninverting terminal. To see how the circuit operates, assume that it is in its stable state with the output at $v_o = +V_Z$ and the capacitor clamped at

$$v_c = V_1 \approx 0.7 \text{ V}$$

(the ON voltage of $D1$ with $\beta V_Z > V_1$). If the trigger amplitude is greater than $\beta V_Z - V_1$, then it will cause the comparator to switch to an output $v_o = -V_Z$. The capacitor will now charge through R' toward $-V_Z$ because $D1$ becomes reverse biased. When v_c becomes more negative than $-\beta V_Z$, the comparator output swings back to $+V_Z$. The capacitor now starts charging toward $+V_Z$ through R' until v_c reaches V_1 and C becomes clamped again at $v_c = V_1$. In Prob. 16-48 we find that the pulse width T is given by

$$T = R'C \ln \frac{1 + (V_1/V_Z)}{1 - \beta} \tag{16-78}$$



(a) (b)

Fig. 16-45 (a) Monostable multivibrator. (b) Output and capacitor voltage waveforms.

If $V_Z \gg V_1$ and $R_2 = R_3$, so that $\beta = \frac{1}{2}$, then $T = 0.69R'C$. For short pulse widths the switching times of the comparator become important and limit the operation of the circuit. If $R_1 = 0$ and the Zener diodes are omitted, Eq. (16-78) remains valid with $V_Z = V_{CC} - V_{CE,\text{sat}}$ (Sec. 15-5).

Triangle-wave Generator We observe from Fig. 16-44*b* that v_c has a triangular waveshape but that the sides of the triangles are exponentials rather than straight lines. To linearize the triangles, it is required that C be charged with a constant current rather than the exponential current supplied through R in Fig. 16-44*b*. In Fig. 16-46 an OP AMP integrator is used to supply constant current to C so that the output is linear. Because of the inversion through the integrator, this voltage is fed back to the noninverting terminal of the comparator in this circuit rather than to the inverting terminal as in Fig. 16-44.

When the comparator has reached either the positive or negative saturation state, the matched Zener diodes will clamp the voltage V_A at either $+V_Z$ or $-V_Z$. Let us assume that $V_A = +V_Z$ at $t = t_o$. The current flowing into the integrator is

$$I^+ = \frac{V_Z}{R_3 + R_4} \tag{16-79}$$

and the integrator output becomes a negative-going ramp, or

$$v_o(t) = v_o(t_o) - \frac{1}{C}\int_{t_0}^{t} I^+\,dt = v_o(t_o) - \frac{I^+}{C}(t - t_o) \tag{16-80}$$

The voltage at pin 3 of the threshold detector is, using superposition,

$$v_3(t) = \frac{R_5 V_Z}{R_1 + R_2 + R_5} + \frac{(R_1 + R_2)v_o(t)}{R_1 + R_2 + R_5} \tag{16-81}$$

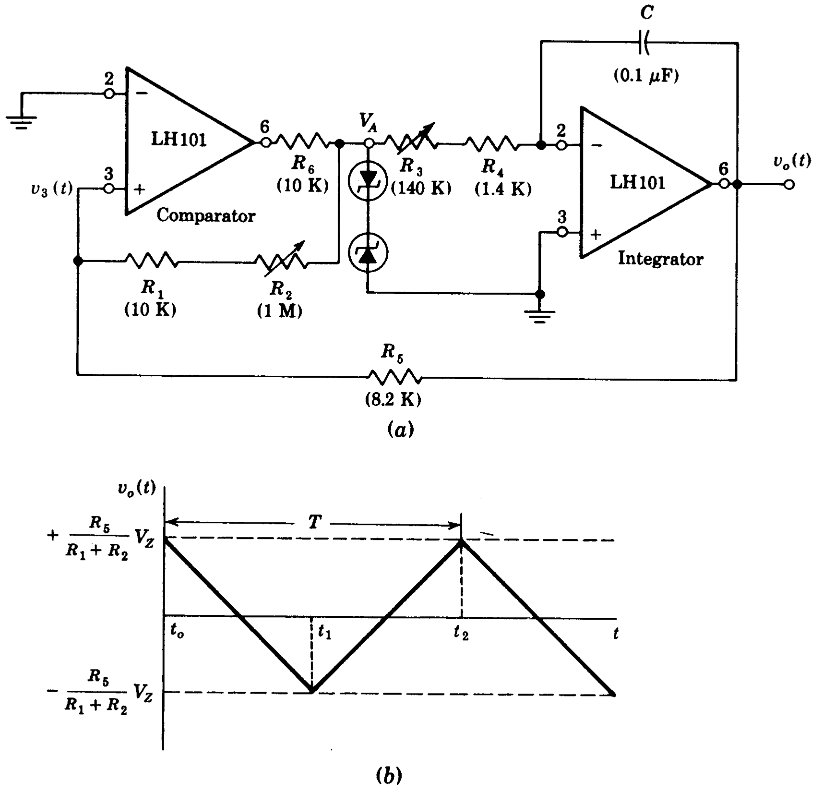

C
(0.1 μF)
2
LH101
6
V_A
R_6
(10 K)
R_3
(140 K)
R_4
(1.4 K)
2
LH101
6
$v_o(t)$
$v_3(t)$
3
Comparator
3
Integrator
R_1
(10 K)
R_2
(1 M)
R_5
(8.2 K)
(a)
$v_o(t)$
$+\frac{R_5}{R_1+R_2}V_Z$
T
t_o
t_1
t_2
t
$-\frac{R_5}{R_1+R_2}V_Z$
(b)

Fig. 16-46 (*a*) Practical triangle-wave generator. (*b*) Output waveform. (Courtesy of National Semiconductor Corp.)

When $v_3(t)$ goes through zero and becomes negative, the comparator output changes to the negative-output state and $V_A = -V_Z$. At this time, $t = t_1$, $v_3(t_1) = 0$, or from Eq. (16-81) we find

$$v_o(t_1) = -\frac{R_5}{R_1 + R_2} V_Z \qquad (16\text{-}82)$$

The current supplied to the integrator for $t_2 > t > t_1$ is

$$I^- = -\frac{V_Z}{R_3 + R_4} = -I^+$$

and the integrator output $v_o(t)$ becomes a positive-going ramp with the same slope as the negative-going ramp. At a time t_2, when

$$v_o(t_2) = +\frac{R_5}{R_1 + R_2} V_Z \qquad (16\text{-}83)$$

the comparator switches again to its positive output and the cycle repeats.

The frequency of the triangle wave is determined from Eq. (16-80) and Fig. (16-46) to be given by

$$f = \frac{R_1 + R_2}{4(R_3 + R_4)R_5C} \tag{16-84}$$

The amplitude can be controlled by the ratio $R_5V_Z/(R_1 + R_2)$. The positive and negative peaks are equal if the Zener diodes are matched. It is possible to offset the triangle with respect to ground if we connect a dc voltage to the inverting terminal of the threshold detector or comparator.

The practical circuit shown in Fig. 16-46 makes use of the LH 101 OP AMP, which is internally compensated for unity-gain feed-back. This monolithic integrated OP AMP has maximum input offset voltage of 5 mV and maximum input bias current of 500 nA. For symmetry of operation the current into the integrator should be large with respect to I_{bias}, and the peak of the output triangle voltage should be large with respect to the input offset voltage.

The design of monostable and astable generators using discrete components is considered in detail in Ref. 3, Chap. 11. One shots constructed from logic gates are indicated in Prob. 17-57.

16-16 REGENERATIVE COMPARATOR (SCHMITT TRIGGER)[14]

As indicated in Fig. 16-33, the transfer characteristic of the MC1530 DIFF AMP makes the change in output from -5 V to $+5$ V for a swing of 2 mV in input voltage. Hence the average slope of this curve or the large-signal voltage gain A_V is $A_V = 10/2 \times 10^{-3} = 5{,}000$. (The incremental gain at the center of the characteristic is calculated in Sec. 15-5 to be 8,670.) By employing positive (regenerative) voltage-series feedback, as is done in Figs. 16-44 and 16-45 for the astable and monostable multivibrators, the gain may be increased greatly. Consequently the total output excursion takes place in a time interval during which the input is changing by much less than 2 mV. Theoretically, if the loop gain $-\beta A_V$ is adjusted to be unity, then the gain with feedback A_{Vf} becomes infinite [Eq. (13-4)]. Such an idealized situation results in an abrupt (zero rise time) transition between the extreme values of output voltage. If a loop gain in excess of unity is chosen, the output waveform continues to be virtually discontinuous at the comparison voltage. However, the circuit now exhibits a phenomenon called *hysteresis*, or *backlash*, which is explained below.

The regenerative comparator of Fig. 16-47*a* is commonly referred to as a *Schmitt trigger* (after the inventor of a vacuum-tube version of this circuit). The input voltage is applied to the inverting terminal 2 and the feedback voltage to the noninverting terminal 1. The feedback factor is $\beta = R_2/(R_1 + R_2)$. For $R_2 = 100\ \Omega$, $R_1 = 10$ K, and $A_V = -5{,}000$, the loop gain is

$$-\beta A_V = 0.1 \times \frac{5{,}000}{10.1} = 49.5 \gg 1$$

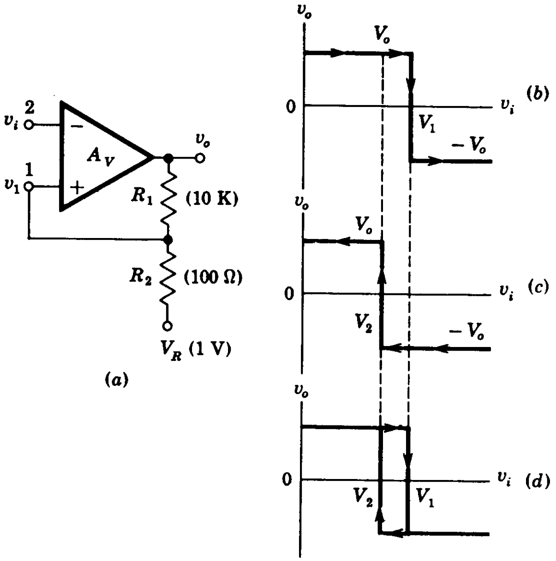


Fig. 16-47 (a) A Schmitt trigger. The transfer characteristics for (b) increasing v_i and (c) decreasing v_i. (d) The compositive input-output curve.

Assume that $v_i < v_1$, so that $v_o = +V_o$ (+5 V). Then, using superposition, we find from Fig. 16-47a that

$$v_1 = \frac{R_1 V_R}{R_1 + R_2} + \frac{R_2 V_o}{R_1 + R_2} \equiv V_1 \tag{16-85}$$

If v_i is now increased, then v_o remains constant at V_o, and $v_1 = V_1 =$ constant until $v_i = V_1$. At this *threshold, critical,* or *triggering voltage,* the output regeneratively switches to $v_o = -V_o$ and remains at this value as long as $v_i > V_1$. This transfer characteristic is indicated in Fig. 16-47b.

The voltage at the noninverting terminal for $v_i > V_1$ is

$$v_1 = \frac{R_1 V_R}{R_1 + R_2} - \frac{R_2 V_o}{R_1 + R_2} \equiv V_2 \tag{16-86}$$

For the parameter values given in Fig. 16-47 and with $V_o = 5\ V$,

$$V_1 = 0.99 + 0.05 = 1.04 \text{ V}$$

$$V_2 = 0.99 - 0.05 = 0.94 \text{ V}$$

Note that $V_2 < V_1$, and the difference between these two values is called the *hysteresis* V_H.

$$V_H = V_1 - V_2 = \frac{2R_2 V_o}{R_1 + R_2} = 0.10 \text{ V} \tag{16-87}$$

If we now decrease v_i, then the output remains at $-V_o$ until v_i equals the voltage at terminal 1 or until $v_i = V_2$. At this voltage a regenerative transition takes place and, as indicated in Fig. 16-47c, the output returns to

$+V_o$ almost instantaneously. The complete transfer function is indicated in Fig. 16-47*c*, where the shaded portions may be traversed in either direction, but the solid segments can only be obtained if v_i varies as indicated by the arrows. Note that because of the hysteresis, the circuit triggers at a higher voltage for increasing than for decreasing signals.

We note above that transfer gain increases from 5,000 toward infinity as the loop gain increases from zero to unity, and that there is no hysteresis as long as $-\beta A_V \leq 1$. However, adjusting the gain precisely to unity is not feasible. The DIFF AMP parameters and, hence the gain A_V, are variable over the signal excursion. Hence an adjustment which ensures that the maximum loop gain is unity would result in voltage ranges where this amplification is less than unity, with a consequent loss in speed of response of the circuit. Furthermore, the circuit may not be stable enough to maintain a loop gain of precisely unity for a long period of time without frequent readjustment. In practice, therefore, a loop gain in excess of unity is chosen and a small amount of hysteresis is tolerated. In most cases a small value of V_H is not a matter of concern. In other applications a large backlash range will not allow the circuit to function properly. Thus if the peak-to-peak signal were smaller than V_H, then the Schmitt circuit, having responded at a threshold voltage by a transition in one direction, would never reset itself. In other words, once the output has jumped to, say, V_o, it would remain at this level and never return to $-V_o$.

The most important use made of the Schmitt trigger is to convert a very slowly varying input voltage into an output having an abrupt (almost discontinuous) waveform, occurring at a precise value of input voltage. This regenerative comparator may be used in all the applications listed in Sec. 16-11. For example, the use of the Schmitt trigger as a squaring circuit is illustrated in Fig. 16-48. The input signal is arbitrary except that it has a large enough excursion to carry the input beyond the limits of the hysteresis range V_H. The output is a square wave as shown, the amplitude of which is independent of the peak-to-peak value of the input waveform. The output has much faster leading and trailing edges than does the input.

The design of a Schmitt trigger from discrete components is explained in detail in Ref. 3, Secs. 10-11 and 10-13.

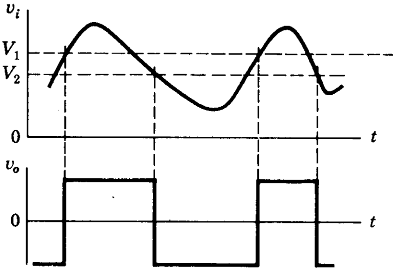


Fig. 16-48 **Response of the Schmitt trigger to an arbitrary input signal.**

16-17 EMITTER-COUPLED LOGIC (ECL)[15]

The transfer characteristic of the difference amplifier is discussed in Sec. 15-4. We find that the emitter current remains essentially constant and that this current is switched from one transistor to the other as the signal at the input transistor varies from about 0.1 V below to 0.1 V above the reference voltage V_{BB} at the base of the second transistor (Fig. 15-9). Except for a very narrow range of input voltage the output voltage takes on only one of two possible values and, hence, behaves as a binary circuit. Hence the DIFF AMF, which is considered in detail in this chapter on analog systems, is also important as a digital device. A logic family based upon this basic building block is called *emitter-coupled logic* (ECL) or *current-mode logic* (CML). Since in the DIFF AMP clipper or comparator neither transistor is allowed to go into saturation, the ECL is the fastest of all logic families (Table 6-5); a propagation delay time as low as 1 ns per gate is possible. The high speed (and high fan out) attainable with ECL is offset by the increased power dissipation per gate relative to that of the saturating logic families.

A 2-input OR (and also NOR) gate is drawn in Fig. 16-49*a*. This circuit is obtained from Fig. 15-6 by using two transistors in parallel at the input. Consider positive logic. If both A and B are low, then neither $Q1$ nor $Q2$ conducts whereas $Q3$ is in its active region. Under these circumstances Y is low and Y' is high. If either A or B is high, then the emitter current switches to the input transistor the base of which is high, and the collector current of $Q3$ drops approximately to zero. Hence Y goes high and Y' drops in voltage. Note that OR logic is performed at the output Y and NOR logic at Y', so that $Y' = \bar{Y}$. The logic symbol for such an OR gate with both true and false outputs is indicated in Fig. 16-49*b*. The availability of complementary out-

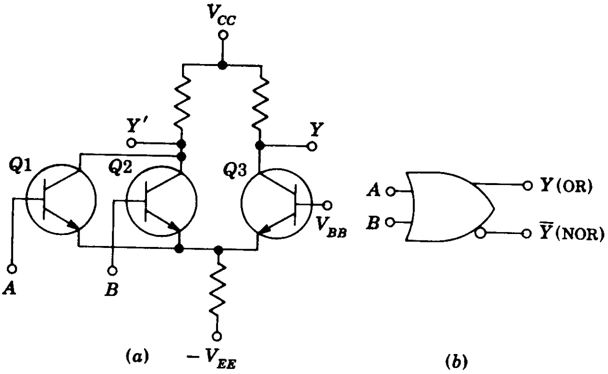


Fig. 16-49 (*a*) DIFF AMP converted into a 2-input emitter-coupled logic circuit. (*b*) The symbol for a 2-input OR/NOR gate.

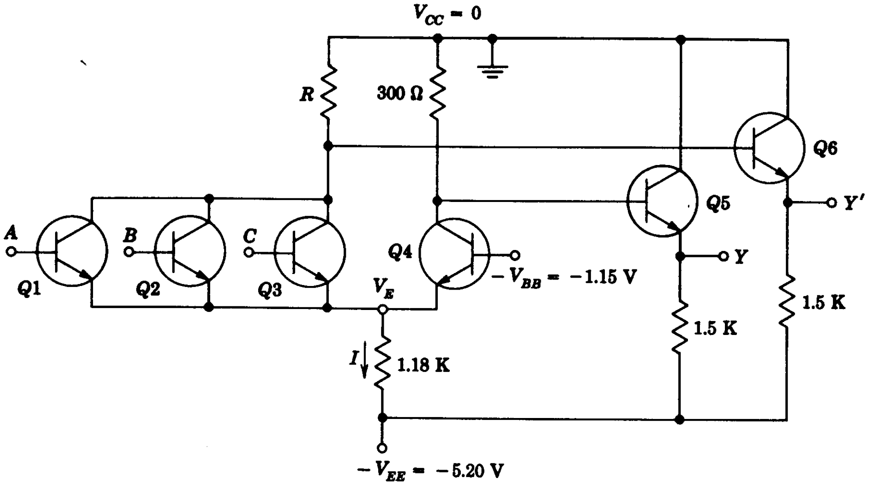


Fig. 16-50 A 3-input ECL OR/NOR gate, with no dc-level shift between input and output voltages.

puts is clearly an advantage to the logic design engineer since it avoids the necessity of adding gates simply as inverters.

One of the difficulties with ECL topology of Fig. 16-49*a* is that the $V(0)$ and $V(1)$ levels at the outputs differ from those at the inputs. Hence emitter followers $Q5$ and $Q6$ are used at the outputs to provide the proper dc-level shifts. The basic Motorola ECL 3-input gate is shown in Fig. 16-50. The reference voltage $-V_{BB}$ is obtained from a temperature-compensated network (not indicated). The quantitative operation of the gate is given in the following illustrative problem.

EXAMPLE (*a*) What are the logic levels at output Y of the ECL gate of Fig. 16-50? Assume a drop of 0.7 V between base and emitter of a conducting transistor. (*b*) Calculate the noise margins. (*c*) Verify that a conducting transistor is in its active region (*not* in saturation). (*d*) Calculate R so that the logic levels at Y' are the complements of those at Y. (*e*) Find the average power dissipated by the gate.

Solution (*a*) If all inputs are low, then assume transistors $Q1$, $Q2$, and $Q3$ are cut off and $Q4$ is conducting. The voltage at the common emitter is

$$V_E = -1.15 - 0.7 = -1.85 \text{ V}$$

The current I in the 1.18-K resistance is

$$I = \frac{-1.85 + 5.20}{1.18} = 2.84 \text{ mA}$$

Neglecting the base current compared with the emitter current, I is the current in the 300-Ω resistance and the output voltage at Y is

$$v_Y = -0.3I - V_{BE5} = -(0.3)(2.84) - 0.7 = -1.55 \text{ V} = V(0)$$

If all inputs are at $V(0) = -1.55$ V and $V_E = -1.85$ V, then the base-to-emitter voltage of an input transistor is

$$V_{BE} = -1.55 + 1.85 = 0.30 \text{ V}$$

Since the cutin voltage is $V_{BE,\text{cutin}} = 0.5$ V (Table 5-1), then the input transistors are nonconducting, as was assumed above.

If at least one input is high, then assume that the current in the 1.18-K resistance is switched to R, and $Q4$ is cut off. The drop in the 300-Ω resistance is then zero. Since the base and collector of $Q5$ are effectively tied together, $Q5$ now behaves as a diode. Assuming 0.7 V across $Q5$ as a first approximation, the diode current is $(5.20 - 0.7)/1.5 = 3.0$ mA. From Fig. 7-19*a* the diode voltage for 3.0 mA is 0.75 V. Hence

$$v_Y = -0.75 \text{ V} = V(1)$$

If one input is at -0.75 V, then $V_E = -0.75 - 0.7 = -1.45$ V, and

$$V_{BE4} = -1.15 + 1.45 = 0.30 \text{ V}$$

which verifies the assumption that $Q4$ is cutoff; since $V_{BE,\text{cutin}} = 0.5$ V.

Note that the total output swing between the two logic levels is only $1.55 - 0.75 = 0.80$ V (800 mV). This voltage is much smaller than the value (in excess of 4 V) obtained with a DTL or TTL gate.

(*b*) If all inputs are at $V(0)$, then the calculation in part (*a*) shows that an input transistor is within $0.50 - 0.30 = 0.20$ V of cutin. Hence a positive noise spike of 0.20 V will cause the gate to malfunction.

If one input is at $V(1)$, then we find in part (*a*) that $V_{BE4} = 0.30$ V. Hence a negative noise spike at the input of 0.20 V drops V_E by the same amount and brings V_{BE4} to 0.5 V, or to the edge of conduction. Note that the noise margins are quite small (±200 mV).

(*c*) From part (*a*) we have that, when $Q4$ is conducting, its collector voltage with respect to ground is the drop in the 300-Ω resistance, or $V_{C4} = -(0.3)(2.84) = -0.85$ V. Hence the collector junction voltage is

$$V_{CB4} = V_{C4} - V_{B4} = -0.85 + 1.15 = +0.30 \text{ V}$$

For an *n-p-n* transistor this represents a reverse bias, and $Q4$ must be in its active region.

If any input, say A, is at $V(1) = -0.75 \text{ V} = V_{B1}$, then $Q1$ is conducting and the output $Y' = \bar{Y} = V(0) = -1.55$ V. The collector of $Q1$ is more positive than $V(0)$ by V_{BE6}, or

$$V_{C1} = -1.55 + 0.7 = -0.85 \text{ V}$$

and

$$V_{CB1} = V_{C1} - V_{B1} = -0.85 + 0.75 = -0.10 \text{ V}$$

For an *n-p-n* transistor this represents a forward bias, but one whose magnitude is less than the cutin voltage of 0.5 V. Therefore $Q1$ is *not* in saturation; it is in its active region.

(*d*) If input A is at $V(1)$, then $Q1$ conducts and $Q4$ is OFF. Then

$$V_E = V(1) - V_{BE1} = -0.75 - 0.7 = -1.45 \text{ V}$$

$$I = \frac{V_E + V_{EE}}{1.18} = \frac{-1.45 + 5.20}{1.18} = 3.17 \text{ mA}$$

In part (*c*) we find that, if $Y' = \bar{Y}$, then $V_{C1} = -0.85$ V. This value represents the drop across R if we neglect the base current of $Q1$. Hence

$$R = \frac{0.85}{3.17} = 0.27 \text{ K} = 270\ \Omega$$

This value of R ensures that, if an input is $V(1)$, then $Y' \doteq V(0)$. If all inputs are at $V(0) = -1.55$ V, then the current through R is zero and the output is $-0.75 \text{ V} = V(1)$, independent of R.

Note that, if I had remained constant as the input changed state (true current-mode switching), then R would be identical to the collector resistance (300 Ω) of $Q4$. The above calculation shows that R is slightly smaller than this value.

(*e*) If the input is low, $I = 2.84$ mA (part *a*), whereas if the input is high, $I = 3.17$ mA (part *d*). The average I is $\frac{1}{2}(2.84 + 3.17) = 3.00$ mA. Since $V(0) = -0.75$ V and $V(1) = -1.55$ V, the currents in the two emitter followers are

$$\frac{5.20 - 0.75}{1.50} = 2.96 \text{ mA} \qquad \text{and} \qquad \frac{5.20 - 1.55}{1.50} = 2.40 \text{ mA}$$

The total power supply current drain is $3.00 + 2.96 + 2.40 = 8.36$ mA and the power dissipation is $(5.20)(8.36) = 43.5$ mW.

Note that the current drain from the power supply varies very little as the input switches from one state to the other. Hence power line spikes (of the type discussed in Sec. 6-12 for TTL gates) are virtually nonexistent.

The input resistance can be considered infinite if all inputs are low so that all input transistors are cut off. If an input is high, then $Q4$ is OFF, and the input resistance corresponds to a transistor with an emitter resistor $R_e = 1.18$ K, and from Eq. (8-55) a reasonable estimate is $R_i \approx 100$ K. The

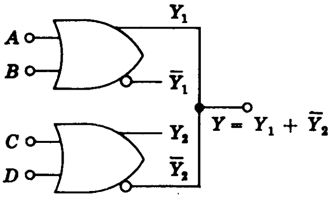


Fig. 16-51 An implied-OR connection at the output of two ECL gates.

output resistance is that of an emitter-follower (or a diode) and a reasonable value is $R_o \approx 15\ \Omega$.

If the outputs of two or more ECL gates are tied together as in Fig. 16-51, then wired-OR logic (Sec. 6-10) is obtained (Prob. 16-53). Open-emitter gates are available for use in this application.

Summary The principal characteristics of the ECL gate are summarized below:

Advantages

1. Since the transistors do not saturate, then the highest speed of any logic family is available.
2. Since the input resistance is very high and the output resistance is very low, a large fan out is possible.
3. Complementary outputs are available.
4. Current switching spikes are not present in the power supply leads.
5. Outputs can be tied together to give the implied-OR function.
6. There is little degradation of parameters with variations in temperature.
7. The number of functions available is high.
8. Easy data transmission over long distances by means of balanced twisted-pair 50-Ω lines is possible.[15]

Disadvantages

1. A small voltage difference (800 mV) exists between the two logic levels and the noise margin is only ±200 mV.
2. The power dissipation is high relative to the other logic families.
3. Level shifters are required for interfacing with other families.
4. The gate is slowed down by heavy capacitive loading.

REFERENCES

1. Korn, G. A., and T. M. Korn: "Electronic Analog and Hybrid Computers," McGraw-Hill Book Company, New York, 1964.
2. Giacoletto, L. J.: "Differential Amplifiers," Wiley-Interscience, New York, 1970.
3. Millman, J., and H. Taub: "Pulse, Digital, and Switching Waveforms," pp. 536–548, McGraw-Hill Book Company, New York, 1965.
4. Huelsman, P. L.: "Active Filters," McGraw-Hill Book Company, New York, 1970.
5. Valley, Jr., C. E., and H. Wallman: "Vacuum Tube Amplifiers," appendix A, McGraw-Hill Book Company, New York, 1948.

6. Kuo, F. F.: "Network Analysis and Synthesis," John Wiley & Sons, Inc., New York, 1962.

7. Stremler, F. G.: "Design of Active Bandpass Filters," *Electronics*, vol. 44, no. 12, pp. 86–89, June 7, 1971.

8. Linvill, J. G., and J. F. Gibbons: "Transistors and Active Circuits," chap. 14, McGraw-Hill Book Company, 1961.

9. Graeme, J. G., and T. E. Tobey: "Operational Amplifiers. Design and Applications," McGraw-Hill Book Company, New York, 1971.

10. Ref. 3, pp. 649–658.

11. Dow, Jr., P. C.: An Analysis of Certain Errors in Electronic Differential Analyzers: Capacitor Dielectric Absorption, *IRE Trans. Electronic Computers*, March, 1958, pp. 17–22.

12. Dobkin, R. C.: "Linear Brief 8," National Semiconductor Corporation, August, 1969.

13. Gilbert, B.: A Precise Four-quadrant Multiplier with Subnanosecond Response, *IEEE J. Solid State Circuits*, December, 1968, p. 210.

14. Sifferlen T. P., and V. Vartanian: "Digital Electronics with Engineering Applications," Prentice-Hall Inc., Englewood Cliffs, N.J. 1970.

15. Garret, L. S.: "Integrated Circuit Digital Logic Families," *IEEE Spectrum*, vol. 7, no. 12, pp. 30–42, December 1970.

REVIEW QUESTIONS

16-1 Indicate an OP AMP connected as (*a*) an *inverter*, (*b*) a *scale changer*, (*c*) a *phase shifter*, and (*d*) an *adder*.

16-2 Draw the circuit of a *voltage-to-current converter* if the load is (*a*) floating and (*b*) grounded.

16-3 Draw the circuit of a *current-to-voltage converter*. Explain its operation.

16-4 Draw the circuit of a dc *voltage follower* and explain its operation.

16-5 Draw the circuit of a dc differential amplifier having (*a*) low input resistance and (*b*) high input resistance.

16-6 Draw the circuit of an ac *voltage follower* having very high input resistance. Explain its operation.

16-7 Draw the circuit of an OP AMP integrator and indicate how to apply the initial condition. Explain its operation.

16-8 Sketch the idealized characteristics for the following filter types: (*a*) low-pass, (*b*) high-pass, (*c*) bandpass, and (*d*) band-rejection.

16-9 Draw the prototype for a low-pass active-filter section of (*a*) first order, (*b*) second order, and (*c*) third order.

16-10 (*a*) Obtain the frequency response of an *RLC* circuit in terms of ω_o and Q. (*b*) Verify that the bandwidth is given by f_o/Q. (*c*) What is meant by an active resonant bandpass filter?

16-11 (*a*) A signal V_s is applied to the inverting terminal (2) of an OP AMP through Z_1 and to the noninverting terminal (1) through Z_2. From (1) to ground is an impedance Z_3, and between (2) and the output is Z_4. Derive the expression for the gain. (*b*) How should Z_1, Z_2, Z_3, and Z_4 be chosen so that the circuit behaves as a *delay equalizer?*

16-12 (*a*) Sketch the basic building block for an IC tuned amplifier. (*b*) Explain how automatic gain control (AGC) is obtained. (*c*) Why does AGC not cause detuning?

16-13 Define the y-parameters (*a*) by equations and (*b*) in words. (*c*) For a cascode circuit which y-parameter is negligible?

16-14 Draw the circuit of an *amplitude modulator* and explain its operation.

16-15 (*a*) Draw the circuit of an IC video amplifier with AGC. (*b*) Sketch the small-signal model.

16-16 (*a*) What does an IC comparator consist of? (*b*) Sketch the transfer characteristic and indicate typical voltage values.

16-17 Sketch the circuit for converting a sinusoid (*a*) into a square wave and (*b*) into a series of positive pulses, one per cycle.

16-18 Explain how to measure the phase difference between two sinusoids.

16-19 Sketch a *sample-and-hold* circuit and explain its operation.

16-20 Sketch the circuit of a precision (*a*) diode and (*b*) clamp and explain their operation.

16-21 (*a*) Sketch the circuit of a fast half-wave rectifier and explain its operation. (*b*) How is this circuit converted into an *average detector?*

16-22 Sketch the circuit of a *peak detector* and explain its operation.

16-23 (*a*) Sketch the circuit of a *logarithmic amplifier* using one OP AMP and explain its operation. (*b*) More complicated logarithmic amplifiers are given in Sec. 16-14. What purpose is served by these circuits?

16-24 In schematic form indicate how to multiply two analog voltages with log-antilog amplifiers.

16-25 Explain how to multiply two analog voltages using a DIFF AMP.

16-26 (*a*) Draw the circuit of a square-wave generator using an OP AMP. (*b*) Explain its operation by drawing the capacitor voltage waveform. (*c*) Derive the expression for the period of a symmetrical waveform.

16-27 (*a*) Draw the circuit of a pulse generator (a monostable multivibrator) using an OP AMP. (*b*) Explain its operation by referring to the capacitor waveform.

16-28 (*a*) Draw the circuit of a triangle generator using a comparator and an integrator. (*b*) Explain its operation by referring to the output waveform. (*c*) What is the peak amplitude?

16-29 (*a*) Sketch a regenerative comparator system and explain its operation. (*b*) What parameters determine the loop gain? (*c*) What parameters determine the hysteresis? (*d*) Sketch the transfer characteristic and indicate the hysteresis.

16-30 (*a*) Sketch a 2-input OR (and also NOR) ECL gate. (*b*) What parameters determine the noise margin? (*c*) Why are the two collector resistors unequal? (*d*) Explain why power line spikes are virtually nonexistent.

16-31 List and discuss at least four advantages and four disadvantages of the ECL gate.

APPENDIX C / PROBLEMS

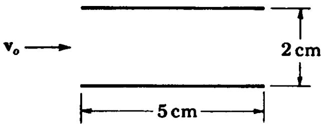
CHAPTER 1

1-1 (*a*) The distance between the plates of a plane-parallel capacitor is 1 cm. An electron starts at rest at the negative plate. If a direct voltage of 1,000 V is applied, how long will it take the electron to reach the positive plate?
(*b*) What is the magnitude of the force which is exerted on the electron at the beginning and at the end of its path?
(*c*) What is its final velocity?
(*d*) If a 60-Hz sinusoidal voltage of peak value 1,000 V is applied, how long will the time of transit be? Assume that the electron is released with zero velocity at the instant of time when the applied voltage is passing through zero. HINT: Expand the sine function into a power series. Thus $\sin\theta = \theta - \theta^3/3! + \theta^5/5! - \cdots$.

1-2 The plates of a parallel-plate capacitor are d m apart. At $t = 0$ an electron is released at the bottom plate with a velocity v_o (meters per second) normal to the plates. The potential of the top plate with respect to the bottom is $-V_m \sin \omega t$.
(*a*) Find the position of the electron at any time t.
(*b*) Find the value of the electric field intensity at the instant when the velocity of the electron is zero.

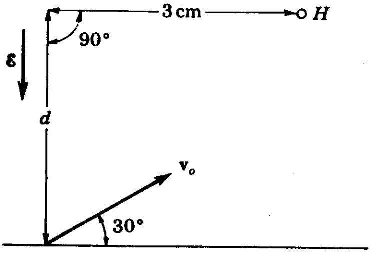
1-3 An electron is released with zero initial velocity from the lower of a pair of horizontal plates which are 3 cm apart. The accelerating potential between these plates increases from zero linearly with time at the rate of 10 V/μs. When the electron is 2.8 cm from the bottom plate, a reverse voltage of 50 V replaces the linearly rising voltage.
(*a*) What is the instantaneous potential between the plates at the time of the potential reversal?
(*b*) With which electrode does the electron collide?
(*c*) What is the time of flight?
(*d*) What is the impact velocity of the electron?

1-4 A 100-eV hydrogen ion is released in the center of the plates, as shown in the figure. The voltage between the plates varies linearly from 0 to 50 V in 10^{-7} s and then drops immediately to zero and remains at zero. The separation between the plates is 2 cm. If the ion enters the region between the plates at time $t = 0$, how far will it be displaced from the X axis upon emergence from between the plates?



Prob. 1-4

1-5 Electrons are projected into the region of constant electric field intensity of magnitude 5×10^3 V/m that exists vertically. The electron-emitting device makes an angle of 30° with the horizontal. It ejects the electrons with an energy of 100 eV.



Prob. 1-5

(*a*) How long does it take an electron leaving the emitting device to pass through a hole H at a horizontal distance of 3 cm from the position of the emitting device? Refer to the figure. Assume that the field is downward.

(*b*) What must be the distance d in order that the particles emerge through the hole?

(*c*) Repeat parts *a* and *b* for the case where the field is upward.

1-6 (*a*) An electron is emitted from an electrode with a negligible initial velocity and is accelerated by a potential of 1,000 V. Calculate the final velocity of the particle.

(*b*) Repeat the problem for the case of a deuterium ion (heavy hydrogen ion—atomic weight 2.01) that has been introduced into the electric field with an initial velocity of 10^5 m/s.

1-7 An electron having an initial kinetic energy of 10^{-16} J at the surface of one of two parallel-plane electrodes and moving normal to the surface is slowed down by the retarding field caused by a 400-V potential applied between the electrodes.

(*a*) Will the electron reach the second electrode?

(*b*) What retarding potential would be required for the electron to reach the second electrode with zero velocity?

1-8 In a certain plane-parallel diode the potential V is given as a function of the distance x between electrodes by the equation

$$V = kx^{4/3}$$

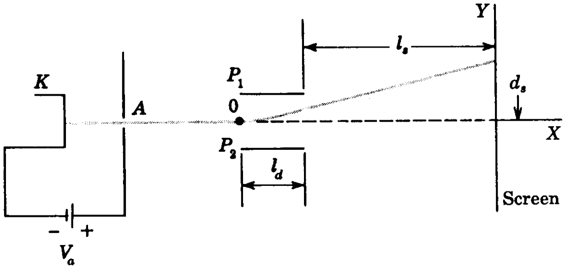
where k is a constant.

(*a*) Find an expression for the time it will take an electron that leaves the electrode with the lower potential with zero initial velocity to reach the electrode with the higher potential, a distance d away.

(*b*) Find an expression for the velocity of this electron.

1-9 The essential features of the displaying tube of an oscilloscope are shown in the accompanying figure. The voltage difference between K and A is V_a and between P_1 and P_2 is V_p. Neither electric field affects the other one. The electrons are emitted from the electrode K with initial zero velocity, and they pass through a hole in the middle of electrode A. Because of the field between P_1 and P_2 they change direction while they pass through these plates and, after that, move with constant velocity toward the screen S. The distance between plates is d.

Prob. 1-9



(*a*) Find the velocity v_x of the electrons as a function of V_a as they cross A.

(*b*) Find the Y-component of velocity v_y of the electrons as they come out of the field of plates P_1 and P_2 as a function of V_p, l_d, d, and v_x.

(*c*) Find the distance from the middle of the screen (d_s), when the electrons reach the screen, as a function of tube distances and applied voltages.

(*d*) For $V_a = 1.0$ kV, and $V_p = 10$ V, $l_d = 1.27$ cm, $d = 0.475$ cm, and $l_s = 19.4$ cm, find the numerical values of v_x, v_y, and d_s.

(*e*) If we want to have a deflection of $d_s = 10$ cm of the electron beam, what must be the value of V_a?

1-10 A diode consists of a plane emitter and a plane-parallel anode separated by a distance of 0.5 cm. The anode is maintained at a potential of 10 V negative with respect to the cathode.

(*a*) If an electron leaves the emitter with a speed of 10^6 m/s, and is directed toward the anode, at what distance from the cathode will it intersect the potential-energy barrier?

(*b*) With what speed must the electron leave the emitter in order to be able to reach the anode?

1-11 A particle when displaced from its equilibrium position is subject to a linear restoring force $f = -kx$, where x is the displacement measured from the equilibrium position. Show by the energy method that the particle will execute periodic vibrations with a maximum displacement which is proportional to the square root of the total energy of the particle.

1-12 A particle of mass m is projected vertically upward in the earth's gravitational field with a speed v_o.

(*a*) Show by the energy method that this particle will reverse its direction at the height of $v_o^2/2g$, where g is the acceleration of gravity.

(*b*) Show that the point of reversal corresponds to a "collision" with the potential-energy barrier.

1-13 (*a*) Prove Eq. (1-13).

(*b*) For the hydrogen atom show that the possible radii in meters are given by

$$r = \frac{h^2 \epsilon_o n^2}{\pi m q^2}$$

where n is any integer but not zero. For the ground state ($n = 1$) show that the radius is 0.53 Å.

1-14 Show that the time for one revolution of the electron in the hydrogen atom in a circular path about the nucleus is

$$T = \frac{m^{\frac{1}{2}} q^2}{4\sqrt{2}\,\epsilon_o(-W)^{\frac{3}{2}}}$$

where the symbols are as defined in Sec. 1-4.

1-15 For the hydrogen atom show that the reciprocal of the wavelength (called the *wave number*) of the spectral lines is given, in waves per meter, by

$$\frac{1}{\lambda} = R\left(\frac{1}{n_2^2} - \frac{1}{n_1^2}\right)$$

where n_1 and n_2 are integers, with n_1 greater than n_2, and $R = mq^4/8\epsilon_o^2 h^3 c = 1.10 \times 10^7\ \text{m}^{-1}$ is called the *Rydberg constant.*

If $n_2 = 1$, this formula gives a series of lines in the ultraviolet, called the *Lyman series.* If $n_2 = 2$, the formula gives a series of lines in the visible, called the *Balmer series.* Similarly, the series for $n_2 = 3$ is called the *Paschen series.* These predicted lines are observed in the hydrogen spectrum.

1-16 Show that Eq. (1-14) is equivalent to Eq. (1-11).

1-17 (*a*) A photon of wavelength 1,026 Å is absorbed by hydrogen, and two other photons are emitted. If one of these is the 1,216 Å line, what is the wavelength of the second photon?

(*b*) If the result of bombardment of the hydrogen was the presence of the fluorescent lines 18,751 and 1,026 Å, what wavelength must have been present in the bombarding radiation?

1-18 The seven lowest energy levels of sodium vapor are 0, 2.10, 3.19, 3.60, 3.75, 4.10, and 4.26 eV. A photon of wavelength 3,300 Å is absorbed by an atom of the vapor.

(*a*) What are all the possible fluorescent lines that may appear?

(*b*) If three photons are emitted and one of these is the 11,380-Å line, what are the wavelengths of the other two photons?

(*c*) Between what energy states do the transitions take place in order to produce these lines?

1-19 (*a*) With what speed must an electron be traveling in a sodium-vapor lamp in order to excite the yellow line whose wavelength is 5,893 Å?

(*b*) What should be the frequency of a photon in order to excite the same yellow line?

(*c*) What would happen if the frequency of the photon was 530 or 490 THz (T = Tera = 10^{12})?

(*d*) What should be the minimum frequency of the photon in order to ionize an unexcited atom of sodium vapor?

(*e*) What should be the minimum speed of an electron in order to ionize an unexcited atom of sodium vapor? Ionization of sodium vapor: 5.12 eV.

1-20 An x-ray tube is essentially a high-voltage diode. The electrons from the hot filament are accelerated by the plate supply voltage so that they fall upon the anode with considerable energy. They are thus able to effect transitions among the tightly bound electrons of the atoms in the solid of which the target (the anode) is constructed.

(*a*) What is the minimum voltage that must be applied across the tube in order to produce x-rays having a wavelength of 0.5 Å?

(*b*) What is the minimum wavelength in the spectrum of an x-ray tube across which is maintained 60 kV?

1-21 Argon resonance radiation corresponding to an energy of 11.6 eV falls upon sodium vapor. If a photon ionizes an unexcited sodium atom, with what speed is the electron ejected? The ionization potential of sodium is 5.12 eV.

1-22 A radio transmitter radiates 1,000 W at a frequency of 10 MHz.

(*a*) What is the energy of each radiated quantum in electron volts?

(*b*) How many quanta are emitted per second?

(*c*) How many quanta are emitted in each period of oscillation of the electromagnetic field?

(*d*) If each quantum acts as a particle, what is its momentum?

1-23 What is the wavelength of (*a*) a mass of 1 kg moving with a speed of 1 m/s, (*b*) an electron which has been accelerated from rest through a potential difference of 10 V?

1-24 Classical physics is valid as long as the physical dimensions of the system are much larger than the De Broglie wavelength. Determine whether the particle is classical in each of the following cases:

(*a*) An electron accelerated through a potential of 300 V in a device whose dimensions are of the order of 1 cm.

(*b*) An electron in the electron beam of a cathode-ray tube (anode-cathode voltage = 25 kV).

(*c*) The electron in a hydrogen atom.

1-25 A photon of wavelength 1,216 Å excites a hydrogen atom which is at rest. Calculate

(*a*) The photon momentum imparted to the atom.

(*b*) The energy corresponding to this momentum and imparted to the hydrogen atom.

(*c*) The ratio of the energy found in part *b* to the energy of the photon. HINT: Use conservation of momentum.

CHAPTER 2

2-1 Prove that the concentration n of free electrons per cubic meter of a metal is given by

$$n = \frac{d\nu}{AM} = \frac{A_o d\nu \times 10^3}{A}$$

where d = density, kg/m^3
ν = valence, free electrons per atom
A = atomic weight
M = weight of atom of unit atomic weight, kg (Appendix A)
A_o = Avogadro's number, molecules/mole

2-2 The specific density of tungsten is 18.8 g/cm^3, and its atomic weight is 184.0. Assume that there are two free electrons per atom. Calculate the concentration of free electrons.

2-3 (*a*) Compute the conductivity of copper for which $\mu = 34.8$ cm^2/V-s and $d = 8.9$ g/cm^3. Use the result of Prob. 2-1.
(*b*) If an electric field is applied across such a copper bar with an intensity of 10 V/cm, find the average velocity of the free electrons.

2-4 Compute the mobility of the free electrons in aluminum for which the density is 2.70 g/cm^3 and the resistivity is 3.44×10^{-6} Ω-cm. Assume that aluminum has three valence electrons per atom. Use the result of Prob. 2-1.

2-5 The resistance of No. 18 copper wire (diameter = 1.03 mm) is 6.51 Ω per 1,000 ft. The concentration of free electrons in copper is 8.4×10^{28} electrons/m^3. If the current is 2 A, find the (*a*) drift velocity, (*b*) mobility, (*c*) conductivity.

2-6 (*a*) Determine the concentration of free electrons and holes in a sample of germanium at 300°K which has a concentration of donor atoms equal to 2×10^{14} atoms/cm^3 and a concentration of acceptor atoms equal to 3×10^{14} atoms/cm^3. Is this *p*- or *n*-type germanium? In other words, is the conductivity due primarily to holes or to electrons?
(*b*) Repeat part *a* for equal donor and acceptor concentrations of 10^{15} atoms/cm^3. Is this *p*- or *n*-type germanium?
(*c*) Repeat part *a* for donor concentration of 10^{16} atoms/cm^3 and acceptor concentration 10^{14} atoms/cm^3.

2-7 (*a*) Find the concentration of holes and of electrons in *p*-type germanium at 300°K if the conductivity is 100 (Ω-cm)$^{-1}$.
(*b*) Repeat part *a* for *n*-type silicon if the conductivity is 0.1 (Ω-cm)$^{-1}$.

2-8 (*a*) Show that the resistivity of intrinsic germanium at 300°K is 45 Ω-cm.
(*b*) If a donor-type impurity is added to the extent of 1 atom per 10^8 germanium atoms, prove that the resistivity drops to 3.7 Ω-cm.

2-9 (*a*) Find the resistivity of intrinsic silicon at 300°K.
(*b*) If a donor-type impurity is added to the extent of 1 atom per 10^8 silicon atoms, find the resistivity.

2-10 Consider intrinsic germanium at room temperature (300°K). By what percent does the conductivity increase per degree rise in temperature?

2-11 Repeat Prob. 2-10 for intrinsic silicon.

2-12 Repeat Prob. 2-6*a* for a temperature of 400°K, and show that the sample is essentially intrinsic.

2-13 A sample of germanium is doped to the extent of 10^{14} donor atoms/cm^3 and 7×10^{13} acceptor atoms/cm^3. At the temperature of the sample the resistivity of pure (intrinsic) germanium is 60 Ω-cm. If the applied electric field is 2 V/cm, find the total conduction current density.

2-14 (*a*) Find the magnitude of the Hall voltage V_H in an *n*-type germanium bar used in Fig. 2-10, having majority-carrier concentration $N_D = 10^{17}/\text{cm}^3$. Assume $B_z = 0.1$ Wb/m^2, $d = 3$ mm, and $\mathcal{E}_x = 5$ V/cm.
(*b*) What happens to V_H if an identical *p*-type germanium bar having $N_A = 10^{17}/\text{cm}^3$ is used in part *a*?

2-15 The Hall effect is used to determine the mobility of holes in a *p*-type silicon bar used in Fig. 2-10. Assume the bar resistivity is 200,000 Ω-cm, the magnetic field $B_z = 0.1$ Wb/m^2, and $d = w = 3$ mm. The measured values of the current and Hall voltage are 10 μA and 50 mV, respectively. Find μ_p.

2-16 A certain photosurface has a spectral sensitivity of 6 mA/W of incident radiation of wavelength 2,537 Å. How many electrons will be emitted photoelectrically by a pulse of radiation consisting of 10,000 photons of this wavelength?

2-17 (*a*) Consider the situation depicted in Fig. 2-13 with the light turned on. Show that the equation of conservation of charge is

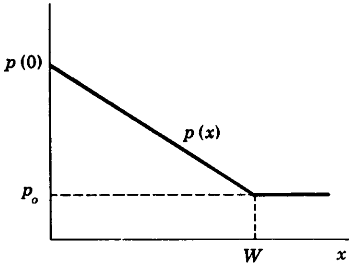
$$\frac{dp}{dt} + \frac{p}{\tau} = \frac{\bar{p}}{\tau}$$

where the time axis in Fig. 2-13 is shifted to t'.
(*b*) Verify that the concentration is given by the equation

$$p = \bar{p} + (p_o - \bar{p})e^{-t/\tau}$$

2-18 The hole concentration in a semiconductor specimen is shown.
(*a*) Find an expression for and sketch the hole current density $J_p(x)$ for the case in which there is no externally applied electric field.

Prob. 2-18



(*b*) Find an expression for and sketch the built-in electric field that must exist if there is to be no net hole current associated with the distribution shown.
(*c*) Find the value of the potential between the points $x = 0$ and $x = W$ if $p(0)/p_o = 10^3$.

2-19 Given a 20 Ω-cm n-type germanium bar with material lifetime of 100 μs, cross section of 1 mm^2, and length of 1 cm. One side of the bar is illuminated with 10^{15} photons/s. Assume that each incident photon generates one electron-hole pair and that these are distributed uniformly throughout the bar. Find the bar resistance under continuous light excitation at room temperature.

2-20 (*a*) Consider an open-circuited graded semiconductor as in Fig. 2-17*a*. Verify the Boltzmann equation for electrons [Eq. (2-61)].
(*b*) For the step-graded semiconductor of Fig. 2-17*b* verify the expression for the contact potential V_o given in Eq. (2-63), starting with $J_n = 0$.

2-21 (*a*) Consider the step-graded germanium semiconductor of Fig. 2-17*b* with $N_D = 10^3 N_A$ and with N_A corresponding to 1 acceptor atom per 10^8 germanium atoms. Calculate the contact difference of potential V_o at room temperature.
(*b*) Repeat part *a* for a silicon p-n junction.

CHAPTER 3

3-1 (*a*) The resistivities of the two sides of a step-graded germanium diode are 2 Ω-cm (p side) and 1 Ω-cm (n side). Calculate the height E_o of the potential-energy barrier.
(*b*) Repeat part *a* for a silicon p-n junction.

3-2 (*a*) Sketch logarithmic and linear plots of carrier concentration vs. distance for an abrupt silicon junction if $N_D = 10^{15}$ atoms/cm^3 and $N_A = 10^{16}$ atoms/cm^3. Give numerical values for ordinates. Label the n, p, and depletion regions.
(*b*) Sketch the space-charge electric field and potential as a function of distance for this case (Fig. 3-1).

3-3 Repeat Prob. 3-2 for an abrupt germanium junction.

3-4 (*a*) Consider a p-n diode operating under low-level injection so that $p_n \ll n_n$. Assuming that the minority current is due entirely to diffusion, verify that the electric field in the n side is given by

$$\mathcal{E}(x) = \frac{I + (D_n/D_p - 1)I_{pn}(x)}{qn\mu_n A}$$

(*b*) Using this value of $\mathcal{E}$, find the next approximation to the drift hole current and show that it may indeed be neglected compared with the diffusion hole current.
(*c*) Sketch the following currents as a function of distance in the n side: (i) total diode current; (ii) minority-carrier current; (iii) majority diffusion current; (iv) majority drift current; (v) total majority-carrier current.

3-5 Starting with Eq. (3-5) for I_{pn} and the corresponding expression for I_{np}, prove that the ratio of hole to electron current crossing a p-n junction is given by

$$\frac{I_{pn}(0)}{I_{np}(0)} = \frac{\sigma_p L_n}{\sigma_n L_p}$$

where $\sigma_p(\sigma_n)$ = conductivity of $p(n)$ side. Note that this ratio depends upon the ratio of the conductivities. For example, if the p side is much more heavily doped than the n side, the hole current will be much larger than the electron current crossing the junction.

3-6 (*a*) Prove that the reverse saturation current in a p-n diode is given by

$$I_o = Aq\left(\frac{D_p}{L_pN_D} + \frac{D_n}{L_nN_A}\right)n_i^2$$

(*b*) Starting with the expression for I_o found in part *a*, verify that the reverse saturation current is given by

$$I_o = AV_T\frac{b\sigma_i^2}{(1+b)^2}\left(\frac{1}{L_p\sigma_n} + \frac{1}{L_n\sigma_p}\right)$$

where $\sigma_n(\sigma_p)$ = conductivity of $n(p)$ side
σ_i = conductivity of intrinsic material
$b = \mu_n/\mu_p$

3-7 (*a*) Using the result of Prob. 3-6, find the reverse saturation current for a germanium p-n junction diode at room temperature, 300°K. The cross-sectional area is 4.0 mm², and

$$\sigma_p = 1.0\ (\Omega\text{-cm})^{-1} \qquad \sigma_n = 0.1\ (\Omega\text{-cm})^{-1} \qquad L_n = L_p = 0.15\text{ cm}$$

Other physical constants are given in Table 2-1.

(*b*) Repeat part *a* for a silicon p-n junction diode. Assume $L_n = L_p = 0.01$ cm and $\sigma_n = \sigma_p = 0.01\ (\Omega\text{-cm})^{-1}$.

3-8 Find the ratio of the reverse saturation current in germanium to that in silicon, using the result of Prob. 3-6. Assume $L_n = L_p = 0.1$ cm and $\sigma_n = \sigma_p = 1.0$ $(\Omega\text{-cm})^{-1}$ for germanium, whereas the corresponding values are 0.01 cm and 0.01 $(\Omega\text{-cm})^{-1}$ for silicon. See also Table 2-1.

3-9 (*a*) For what voltage will the reverse current in a p-n junction germanium diode reach 90 percent of its saturation value at room temperature?

(*b*) What is the ratio of the current for a forward bias of 0.05 V to the current for the same magnitude of reverse bias?

(*c*) If the reverse saturation current is 10 μA, calculate the forward currents for voltages of 0.1, 0.2, and 0.3 V, respectively.

3-10 (*a*) Evaluate η in Eq. (3-9) from the slope of the plot in Fig. 3-8 for $T = 25$°C. Draw the best-fit line over the current range 0.01 to 10 mA.

(*b*) Repeat for $T = -55$ and 150°C.

3-11 (*a*) Calculate the anticipated factor by which the reverse saturation current of a germanium diode is multiplied when the temperature is increased from 25 to 80°C.

(*b*) Repeat part *a* for a silicon diode over the range 25 to 150°C.

3-12 It is predicted that, for germanium, the reverse saturation current should increase by 0.11°C^{-1}. It is found experimentally in a particular diode that at a reverse voltage of 10 V, the reverse current is 5 μA and the temperature dependence is only 0.07°C^{-1}. What is the leakage resistance shunting the diode?

3-13 A diode is mounted on a chassis in such a manner that, for each degree of temperature rise above ambient, 0.1 mW is thermally transferred from the diode to its surroundings. (The "thermal resistance" of the mechanical contact between the diode and its surroundings is 0.1 mW/°C.) The ambient temperature is 25°C. The diode temperature is not to be allowed to increase by more than 10°C above ambient. If the reverse saturation current is 5.0 μA at 25°C and increases at the rate 0.07°C^{-1}, what is the maximum reverse-bias voltage which may be maintained across the diode?

3-14 A silicon diode operates at a forward voltage of 0.4 V. Calculate the factor by which the current will be multiplied when the temperature is increased from 25 to 150°C. Compare the result with the plot of Fig. 3-8.

3-15 An ideal germanium *p-n* junction diode has at a temperature of 125°C a reverse saturation current of 30 μA. At a temperature of 125°C find the dynamic resistance for a 0.2 V bias in (*a*) the forward direction, (*b*) the reverse direction.

3-16 Prove that for an alloy *p-n* junction (with $N_A \ll N_D$), the width W of the depletion layer is given by

$$W = \left(\frac{2\epsilon\mu_p V_j}{\sigma_p}\right)^{\frac{1}{2}}$$

where V_j is the junction potential under the condition of an applied diode voltage V_d.

3-17 (*a*) Prove that for an alloy silicon *p-n* junction (with $N_A \ll N_D$), the depletion-layer capacitance in picofarads per square centimeter is given by

$$C_T = 2.9 \times 10^{-4}\left(\frac{N_A}{V_j}\right)^{\frac{1}{2}}$$

(*b*) If the resistivity of the *p* material is 3.5 Ω-cm, the barrier height V_o is 0.35 V, the applied reverse voltage is 5 V, and the cross-sectional area is circular of 40 mils diameter, find C_T.

3-18 (*a*) For the junction of Fig. 3-10, find the expression for the $\mathcal{E}$ and V as a function of x in the *n*-type side for the case where N_A and N_D are of comparable magnitude. HINT: Shift the origin of x so that $x = 0$ at the junction.
(*b*) Show that the total barrier voltage is given by Eq. (3-21) multiplied by $N_A/(N_A + N_D)$ and with $W = W_p + W_n$.
(*c*) Prove that $C_T = [qN_AN_D\epsilon/2(N_A + N_D)]^{\frac{1}{2}}V^{-\frac{1}{2}}$.
(*d*) Prove that $C_T = \epsilon A/(W_p + W_n)$.

3-19 Reverse-biased diodes are frequently employed as electrically controllable variable capacitors. The transition capacitance of an abrupt junction diode is 20 pF at 5 V. Compute the decrease in capacitance for a 1.0-V increase in bias.

3-20 Calculate the barrier capacitance of a germanium *p-n* junction whose area is 1 mm by 1 mm and whose space-charge thickness is 2×10^{-4} cm. The dielectric constant of germanium (relative to free space) is 16.

3-21 The zero-voltage barrier height at an alloy-germanium *p-n* junction is 0.2 V. The concentration N_A of acceptor atoms in the *p* side is much smaller than the concentration of donor atoms in the *n* material, and $N_A = 3 \times 10^{20}$ atoms/m^3. Calculate the width of the depletion layer for an applied reverse voltage of (*a*) 10 V and (*b*) 0.1 V and (*c*) for a forward bias of 0.1 V. (*d*) If the cross-sectional area of the diode is 1 mm², evaluate the space-charge capacitance corresponding to the values of applied voltage in (*a*) and (*b*).

3-22 (*a*) Consider a grown junction for which the uncovered charge density ρ varies linearly with distance. If $\rho = ax$, prove that the barrier voltage V_j is given by

$$V_j = \frac{aW^3}{12\epsilon}$$

(*b*) Verify that the barrier capacitance C_T is given by Eq. (3-23)

3-23 Given a forward-biased silicon diode with $I = 1$ mA. If the diffusion capacitance is $C_D = 1\ \mu$F, what is the diffusion length L_p? Assume that the doping of the *p* side is much greater than that of the *n* side.

3-24 The derivation of Eq. (3-28) for the diffusion capacitance assumes that the *p* side is much more heavily doped than the *n* side, so that the current at the junction is entirely due to holes. Derive an expression for the total diffusion capacitance when this approximation is not made.

3-25 (*a*) Prove that the maximum electric field $\mathcal{E}_m$ at a step-graded junction with $N_A \gg N_D$ is given by

$$\mathcal{E}_m = \frac{2V_j}{W}$$

(*b*) It is found that Zener breakdown occurs when $\mathcal{E}_m = 2 \times 10^7$ V/*m* $\equiv \mathcal{E}_z$. Prove that Zener voltage V_z is given by

$$V_z = \frac{\epsilon\mathcal{E}_z^2}{2qN_D}$$

Note that the Zener breakdown voltage can be controlled by controlling the concentration of donor ions.

3-26 (*a*) Zener breakdown occurs in germanium at a field intensity of 2×10^7 V/m. Prove that the breakdown voltage is $V_Z = 51/\sigma_p$, where σ_p is the conductivity of the *p* material in $(\Omega\text{-cm})^{-1}$. Assume that $N_A \ll N_D$.
(*b*) If the *p* material is essentially intrinsic, calculate V_Z.
(*c*) For a doping of 1 part in 10^8 of *p*-type material, the resistivity drops to 3.7 Ω-cm. Calculate V_Z.
(*d*) For what resistivity of the *p*-type material will $V_Z = 1$ V?

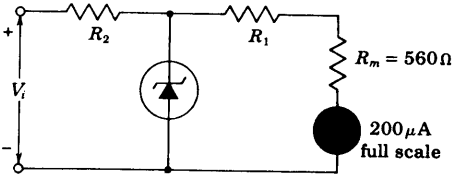
3-27 (*a*) Two *p-n* germanium diodes are connected in series opposing. A 5-V battery is impressed upon this series arrangement. Find the voltage across each junction at room temperature. Assume that the magnitude of the Zener voltage is greater than 5 V.

Note that the result is independent of the reverse saturation current. Is it also independent of temperature?

HINT: Assume that reverse saturation current flows in the circuit, and then justify this assumption.

(*b*) If the magnitude of the Zener voltage is 4.9 V, what will be the current in the circuit? The reverse saturation current is 5 μA.

3-28 The Zener diode can be used to prevent overloading of sensitive meter movements without affecting meter linearity. The circuit shown represents a dc voltmeter which reads 20 V full scale. The meter resistance is 560 Ω, and $R_1 + R_2 = 99.5$ K. If the diode is a 16-V Zener, find R_1 and R_2 so that, when $V_i > 20$ V, the Zener diode conducts and the overload current is shunted away from the meter.



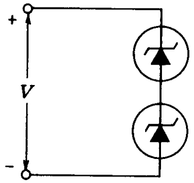
Prob. 3-28

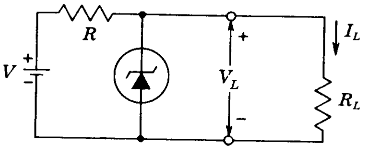
3-29 A series combination of a 15-V avalanche diode and a forward-biased silicon diode is to be used to construct a zero-temperature-coefficient voltage reference. The temperature coefficient of the silicon diode is −1.7 mV/°C. Express in percent per degree centigrade the required temperature coefficient of the Zener diode.

3-30 The saturation currents of the two diodes are 1 and 2 μA. The breakdown voltages of the diodes are the same and are equal to 100 V.

(*a*) Calculate the current and voltage for each diode if $V = 90$ V and $V = 110$ V.

(*b*) Repeat part *a* if each diode is shunted by a 10-M resistor.





Prob. 3-30

Prob. 3-31

3-31 (*a*) The avalanche diode regulates at 50 V over a range of diode currents from 5 to 40 mA. The supply voltage $V = 200$ V. Calculate R to allow voltage regulation from a load current $I_L = 0$ up to I_{max}, the maximum possible value of I_L. What is I_{max}?

(*b*) If R is set as in part *a* and the load current is set at $I_L = 25$ mA, what are the limits between which V may vary without loss of regulation in the circuit?

3-32 (*a*) Consider a tunnel diode with $N_D = N_A$ and with the impurity concentration corresponding to 1 atom per 10^3 germanium atoms. At room temperature calculate (i) the height of the potential-energy barrier under open-circuit conditions (the contact potential energy), (ii) the width of the space-charge region.

(*b*) Repeat part *a* if the semiconductor is silicon instead of germanium.

3-33 The photocurrent I in a *p-n* junction photodiode as a function of the distance x of the light spot from the junction is given in Fig. 3-22. Prove that the slopes of ln I versus x are $-1/L_p$ and $-1/L_n$, respectively, on the n and p sides. Note that L_p represents the diffusion length for holes in the n material.

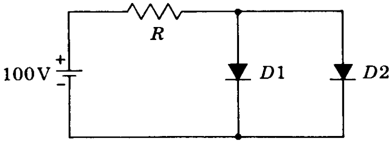
3-34 (*a*) For the type LS 223 photovoltaic cell whose characteristics are given in Fig. 3-23, plot the power output vs. the load resistance R_L.
(*b*) What is the optimum value of R_L?

CHAPTER 4

4-1 (*a*) In the circuit of Prob. 3-27, the Zener breakdown voltage is 2.0 V. The reverse saturation current is 5 μA. If the silicon diode resistance could be neglected, what would be the current?
(*b*) If the ohmic resistance is 100 Ω, what is the current?
Note: Answer part *b* by plotting Eq. (3-9) and drawing a load line. Verify your answer analytically by a method of successive approximations.

4-2 A *p-n* germanium junction diode at room temperature has a reverse saturation current of 10 μA, negligible ohmic resistance, and a Zener breakdown voltage of 100 V. A 1-K resistor is in series with this diode, and a 30-V battery is impressed across this combination. Find the current (*a*) if the diode is forward-biased, (*b*) if the battery is inserted into the circuit with the reverse polarity. (*c*) Repeat parts *a* and *b* if the Zener breakdown voltage is 10 V.

4-3 Each diode is described by a linearized volt-ampere characteristic, with incremental resistance r and offset voltage V_γ. Diode $D1$ is germanium with V_γ = 0.2 V and r = 20 Ω, whereas $D2$ is silicon with V_γ = 0.6 V and r = 15 Ω. Find the diode currents if (*a*) R = 10 K, (*b*) R = 1 K.



Prob. 4-3

4-4 The photodiode whose characteristics are given in Fig. 3-21 is in series with a 30-V supply and a resistance R. If the illumination is 3,000 fc, find the current for (*a*) R = 0, (*b*) R = 50 K, (*c*) R = 100 K.

4-5 (*a*) For the application in Sec. 4-3, plot the voltage across the diode for one cycle of the input voltage v_i. Let V_m = 2.4 V, V_γ = 0.6 V, R_f = 10 Ω, and R_L = 100 Ω.
(*b*) By direct integration find the average value of the diode voltage and the load voltage. Note that these two answers are numerically equal and explain why.

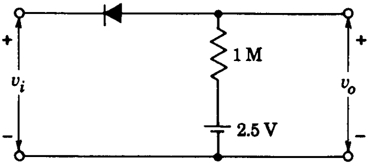
4-6 Calculate the break region over which the dynamic resistance of a diode is multiplied by a factor of 1,000.

4-7 For the diode clipping circuit of Fig. 4-9*a* assume that V_R = 10 V, v_i = 20 sin ωt, and that the diode forward resistance is R_f = 100 Ω while R_r = ∞ and V_γ = 0. Neglect all capacitances. Draw to scale the input and output waveforms and label the maximum and minimum values if (*a*) $R = 100\ \Omega$, (*b*) $R = 1$ K, and (*c*) $R = 10$ K.

4-8 Repeat Prob. 4-7 for the case where the reverse resistance is $R_r = 10$ K.

4-9 In the diode clipping circuit of Fig. 4-9*a* and *d*, $v_i = 20 \sin \omega t$, $R = 1$ K, and $V_R = 10$ V. The reference voltage is obtained from a tap on a 10-K divider connected to a 100-V source. Neglect all capacitances. The diode forward resistance is $50\ \Omega$, $R_r = \infty$, and $V_\gamma = 0$. In both cases draw the input and output waveforms to scale. Which circuit is the better clipper? HINT: Apply Thévenin's theorem to the reference-voltage divider network.

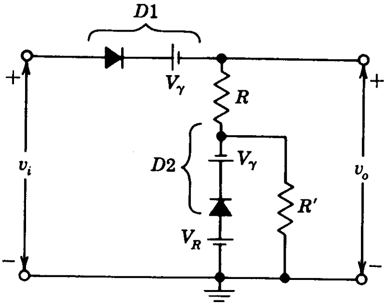
4-10 A symmetrical 5-kHz square wave whose output varies between +10 and −10 V is impressed upon the clipping circuit shown. Assume $R_f = 0$, $R_r = 2$ M, and $V_\gamma = 0$. Sketch the steady-state output waveform, indicating numerical values of the maximum, minimum, and constant portions.



Prob. 4-10

4-11 For the clipping circuits shown in Fig. 4-9*b* and *d* derive the transfer characteristic v_o versus v_i, taking into account R_f and V_γ and considering $R_r = \infty$.

4-12 The clipping circuit shown employs temperature compensation. The dc voltage source V_γ represents the diode offset voltage; otherwise the diodes are assumed to be ideal with $R_f = 0$ and $R_r = \infty$.



Prob. 4-12

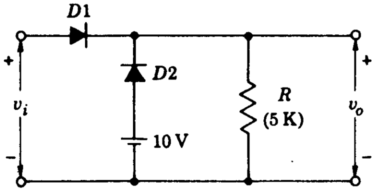
(*a*) Sketch the transfer curve v_o versus v_i.

(*b*) Show that the maximum value of the input voltage v_i so that the current in *D*2 is always in the forward direction is

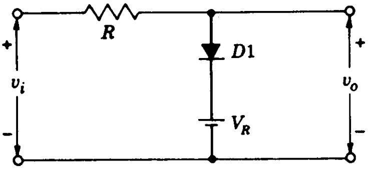
$$v_{i,\max} = V_R + \frac{R}{R'}(V_R - V_\gamma)$$

(*c*) What is the temperature dependence of the point on the input waveform at which clipping occurs?

4-13 (*a*) In the clipping circuit shown, $D2$ compensates for temperature variations. Assume that the diodes have infinite back resistance, a forward resistance of 50 Ω, and a break point at the origin ($V_\gamma = 0$). Calculate and plot the transfer characteristic v_o against v_i. Show that the circuit has an extended break point, that is, two break points close together.



Prob. 4-13



Prob. 4-14

(*b*) Find the transfer characteristic that would result if $D2$ were removed and the resistor R were moved to replace $D2$.

(*c*) Show that the double break of part *a* would vanish and only the single break of part *b* would appear if the diode forward resistances were made vanishingly small in comparison with R.

4-14 (*a*) In the peak clipping circuit shown, add another diode $D2$ and a resistor R' in a manner that will compensate for drift with temperature.

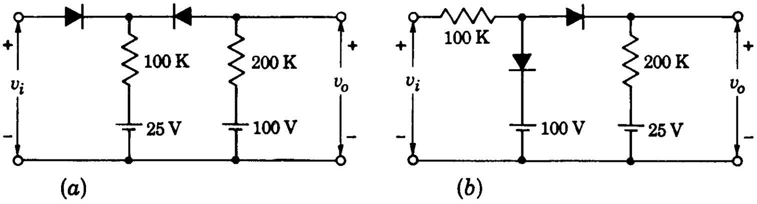
(*b*) Show that the break point of the transmission curve occurs at V_R. Assume $R_r \gg R \gg R_f$.

(*c*) Show that if $D2$ is always to remain in conduction it is necessary that

$$v_i < v_{i,\max} = V_R + \frac{R}{R'}(V_R - V_\gamma)$$

4-15 (*a*) The input voltage v_i to the two-level clipper shown in part *a* of the figure varies linearly from 0 to 150 V. Sketch the output voltage v_o to the same time scale as the input voltage. Assume ideal diodes.

(*b*) Repeat (*a*) for the circuit shown in part *b* of the figure.



Prob. 4-15

4-16 The circuit of Fig. 4-10*a* is used to "square" a 10-kHz input sine wave whose peak value is 50 V. It is desired that the output voltage waveform be flat for 90 percent of the time. Diodes are used having a forward resistance of 100 Ω and a backward resistance of 100 K.

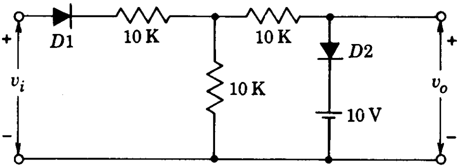
(*a*) Find the values of V_{R1} and V_{R2}.

(*b*) What is a reasonable value to use for R?

4-17 (*a*) The diodes are ideal. Write the transfer characteristic equations (v_o as a function of v_i).

(*b*) Plot v_o against v_i, indicating all intercepts, slopes, and voltage levels.

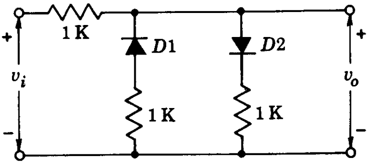
(*c*) Sketch v_o if $v_i = 40 \sin \omega t$. Indicate all voltage levels.



Prob. 4-17

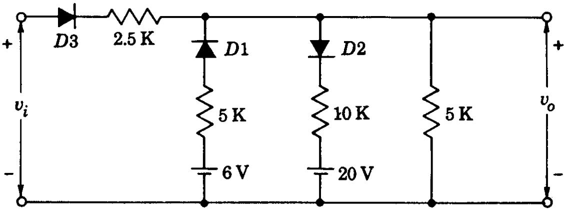
4-18 (*a*) Repeat Prob. 4-17 for the circuit shown.

(*b*) Repeat for the case where the diodes have an offset voltage $V_\gamma = 1$ V.



Prob. 4-18

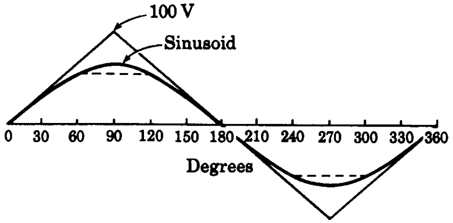
4-19 Assume that the diodes are ideal. Make a plot of v_o against v_i for the range of v_i from 0 to 50 V. Indicate all slopes and voltage levels. Indicate, for each region, which diodes are conducting.



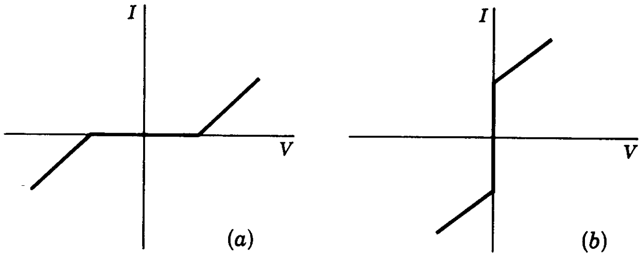
Prob. 4-19

4-20 The triangular waveform shown is to be converted into a sine wave by using clipping diodes. Consider the dashed waveform sketched as a first approximation to the sinusoid. The dashed waveform is coincident with the sinusoid at 0°, 30°, 60°, etc: Devise a circuit whose output is this broken-line waveform when the input is the triangular waveform. Assume ideal diodes and calculate the values of all supply voltages and resistances used. The peak value of the sinusoid is 50 V.

Prob. 4-20



4-21 Construct circuits which exhibit terminal characteristics as shown in parts *a* and *b* of the figure.



Prob. 4-21

4-22 The diode-resistor comparator of Fig. 4-13 is connected to a device which responds when the comparator output attains a level of 0.1 V. The input is a ramp which rises at the rate 10 V/μs. The germanium diode has a reverse saturation current of 1 μA. Initially, $R = 1$ K and the 0.1-V output level is attained at a time $t = t_1$. If we now set $R = 100$ K, what will be the corresponding change in t_1? $V_R = 0$.

4-23 For the four-diode sampling gate of Fig. 4-14 consider that v_s is at its most negative value, say, $v_s = -V_s$. Then verify that the expressions for $V_{n,\min}$ and $V_{c,\min}$ given in Eqs. (4-7) and (4-8) remain valid.

4-24 A balancing voltage divider is inserted between $D3$ and $D4$ in Fig. 4-14 so as to give zero output voltage for zero input. If the divider is assumed to be set at its midpoint, if its total resistance is R, and if R and R_f are both much less than R_c or R_L, show that

$$V_{c,\min} = V_s\left(2 + \frac{R_c}{R_L}\right)\left(1 + \frac{R}{4R_f}\right)$$

4-25 (*a*) Explain qualitatively the operation of the sampling gate shown. The supply voltage V is constant. The control voltage v_c is the square wave of Fig. 4-14*b*. Assume ideal diodes with $V_\gamma = 0$, $R_f = 0$, and $R_r = \infty$. HINT: When $v_c = V_c$, the diodes $D1$ and $D2$ conduct (if $V > V_{\text{min}}$) and $D3$ and $D4$ are OFF. If $v_c = -V_n$, then $D3$ and $D4$ conduct and $D1$ and $D2$ are OFF.

Verify the following relationships:

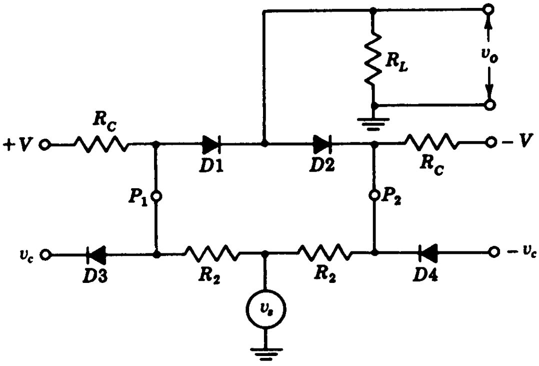
(*b*) $$V_{\text{min}} = \frac{R_c}{R_2}\,\frac{R_1}{R_1 + 2R_L}\,V_s$$

where $R_1 = R_c \| R_2$.

(*c*) $$A = \frac{v_o}{v_s} = \frac{2R_L}{R_2}\,\frac{R_1}{R_1 + 2R_L}$$

(*d*) $$V_{n,\text{min}} = V_s\,\frac{R_c}{R_c + R_2} - V\,\frac{R_2}{R_c + R_2}$$

(*e*) $$V_{c,\text{min}} = A V_s$$



Prob. 4-25

4-26 A diode whose internal resistance is 20 Ω is to supply power to a 1,000-Ω load from a 110-V (rms) source of supply. Calculate (*a*) the peak load current, (*b*) the dc load current, (*c*) the ac load current, (*d*) the dc diode voltage, (*e*) the total input power to the circuit, (*f*) the percentage regulation from no load to the given load.

4-27 Show that the maximum dc output power $P_{\text{dc}} \equiv V_{\text{dc}} I_{\text{dc}}$ in a half-wave single-phase circuit occurs when the load resistance equals the diode resistance R_f.

4-28 The efficiency of rectification η_r is defined as the ratio of the dc output power $P_{\text{dc}} \equiv V_{\text{dc}} I_{\text{dc}}$ to the input power $P_i = (1/2\pi) \int_0^{2\pi} v_i i \, d\alpha$.

(*a*) Show that, for the half-wave-rectifier circuit,

$$\eta_r = \frac{40.6}{1 + R_f/R_L}\ \%$$

(*b*) Show that, for the full-wave rectifier, η_r has twice the value given in part *a*.

4-29 Prove that the regulation of both the half-wave and the full-wave rectifier is given by

$$\%\ \text{regulation} = \frac{R_f}{R_L} = 100\%$$

4-30 (*a*) Prove Eqs. (4-21) and (4-22) for the dc voltage of a full-wave-rectifier circuit. (*b*) Find the dc voltage across a diode by direct integration.

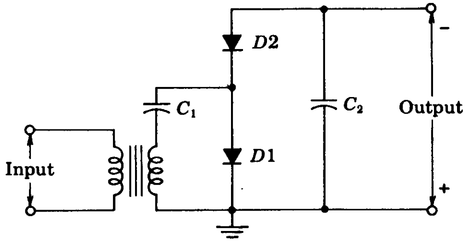
4-31 A full-wave single-phase rectifier consists of a double-diode vacuum tube, the internal resistance of each element of which may be considered to be constant and equal to 500 Ω. These feed into a pure resistance load of 2,000 Ω. The secondary transformer voltage to center tap is 280 V. Calculate (*a*) the dc load current, (*b*) the direct current in each tube, (*c*) the ac voltage across each diode, (*d*) the dc output power, (*e*) the percentage regulation.

4-32 In the full-wave single-phase bridge, can the transformer and the load be interchanged? Explain carefully.

4-33 A 1-mA dc meter whose resistance is 10 Ω is calibrated to read rms volts when used in a bridge circuit with semiconductor diodes. The effective resistance of each element may be considered to be zero in the forward direction and infinite in the inverse direction. The sinusoidal input voltage is applied in series with a 5-K resistance. What is the full-scale reading of this meter?

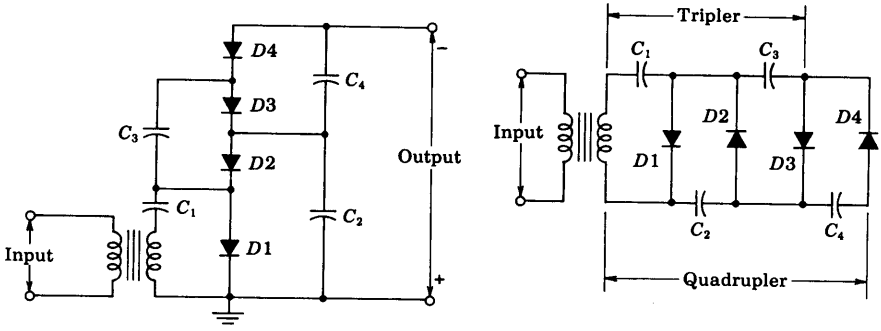
4-34 The circuit shown is a half-wave voltage doubler. Analyze the operation of this circuit. Calculate (*a*) the maximum possible voltage across each capacitor, (*b*) the peak inverse voltage of each diode. Compare this circuit with the bridge voltage doubler of Fig. 4-22. In this circuit the output voltage is negative with respect to ground. Show that if the connections to the cathode and anode of each diode are interchanged, the output voltage will be positive with respect to ground.

Prob. 4-34



4-35 The circuit of Prob. 4-34 can be extended from a doubler to a quadrupler by adding two diodes and two capacitors as shown. In the figure, parts *a* and *b* are alternative ways of drawing the same circuit.

(*a*) Analyze the operation of this circuit.
(*b*) Answer the same questions as asked in Prob. 4-34.
(*c*) Generalize the circuit of this and of Prob. 4-34 so as to obtain *n*-fold multiplication when *n* is any even number. In particular, sketch the circuit for sixfold multiplication.
(*d*) Show that *n*-fold multiplication, with *n* odd, can also be obtained provided that the output is properly chosen.



(*a*) (*b*)

Prob. 4-35

4-36 A single-phase full-wave rectifier uses a semiconductor diode. The transformer voltage is 35 V rms to center tap. The load consists of a 40-μF capacitance in parallel with a 250-Ω resistor. The diode and the transformer resistances and leakage reactance may be neglected.
(*a*) Calculate the cutout angle.
(*b*) Plot to scale the output voltage and the diode current as in Fig. 4-25. Determine the cutin point graphically from this plot, and find the peak diode current corresponding to this point.
(*c*) Repeat parts *a* and *b*, using a 160-μF instead of a 40-μF capacitance.

CHAPTER 5

5-1 (*a*) Show that for an *n-p-n* silicon transistor of the alloy type in which the resistivity ρ_B of the base is much larger than that of the collector, the punch-through voltage V is given by $V = 6.1 \times 10^3 W_B^2/\rho_B$, where V is in volts, ρ_B in ohm-centimeters, and W in mils. For punch-through, $W = W_B$ in Fig. 5-8*a*.
(*b*) Calculate the punch-through voltage if $W = 1$ μm and $\rho_B = 0.5$ Ω-cm.

5-2 The transistor of Fig. 5-3*a* has the characteristics given in Figs. 5-6 and 5-7. Let $V_{CC} = 6$ V, $R_L = 200$ Ω, and $I_E = 15$ mA.
(*a*) Find I_C and V_{CB}.

(*b*) Find V_{EB} and V_L.

(*c*) If I_E changes by $\Delta I_E = 10$ mA symmetrically around the point of part *a* and with constant V_{CB}, find the corresponding change in I_C.

5-3 The CB transistor used in the circuit of Fig. 5-3*a* has the characteristics given in Figs. 5-6 and 5-7. Let $I_C = -20$ mA, $V_{CB} = -4$ V, and $R_L = 200\ \Omega$.

(*a*) Find V_{CC} and I_E.

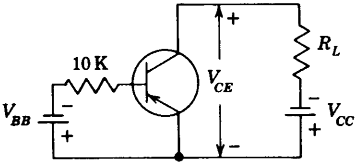
(*b*) If the supply voltage V_{CC} decreases from its value in part *a* by 2 V while I_E retains its previous value, find the new values of I_C and V_{CB}.

5-4 The CE transistor used in the circuit shown has the characteristics given in Figs. 5-10 and 5-11.

(*a*) Find V_{BB} if $V_{CC} = 10$ V, $V_{CE} = -1$ V, and $R_L = 250\ \Omega$.

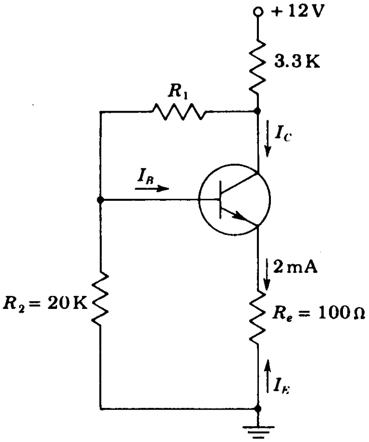
(*b*) If $V_{CC} = 10$ V, find R_L so that $I_C = -20$ mA and $V_{CE} = -4$ V. Find V_{BB}.

Prob. 5-4



5-5 If $\alpha = 0.98$ and $V_{BE} = 0.7$ V, find R_1 in the circuit shown for an emitter current $I_E = -2$ mA. Neglect the reverse saturation current.

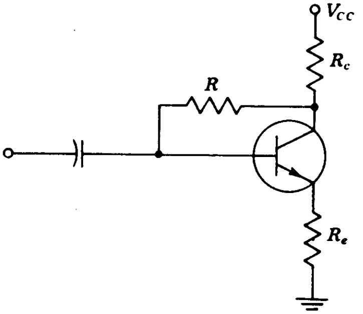
Prob. 5-5



5-6 (*a*) Find R_c and R_b in the circuit of Fig. 5-12*a* if $V_{CC} = 10$ V and $V_{BB} = 5$ V, so that $I_C = 10$ mA and $V_{CE} = 5$ V. A silicon transistor with $\beta = 100$, $V_{BE} = 0.7$ V, and negligible reverse saturation current is under consideration.

(*b*) Repeat part *a* if a 100-Ω emitter resistor is added to the circuit.

5-7 In the circuit shown, $V_{CC} = 24$ V, $R_c = 10$ K, and $R_e = 270\ \Omega$. If a silicon transistor is used with $\beta = 45$ and if $V_{CE} = 5$ V, find R. Neglect the reverse saturation current.

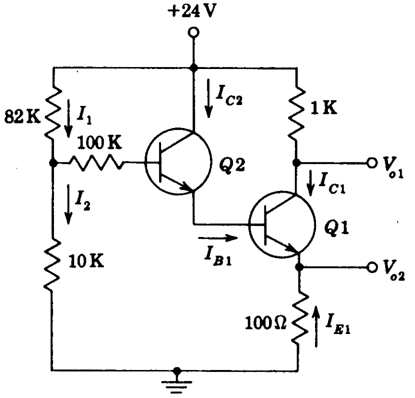


Prob. 5-7

5-8 For the circuit shown, transistors $Q1$ and $Q2$ operate in the active region with $V_{BE1} = V_{BE2} = 0.7$ V, $\beta_1 = 100$, and $\beta_2 = 50$. The reverse saturation currents may be neglected.

(*a*) Find the currents I_{B2}, I_1, I_2, I_{C2}, I_{B1}, I_{C1}, and I_{E1}.

(*b*) Find the voltages V_{o1} and V_{o2}.

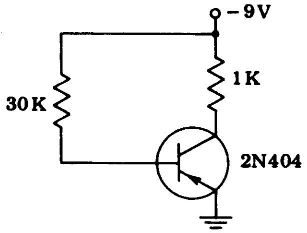


Prob. 5-8

5-9 (*a*) The reverse saturation current of the germanium transistor in Fig. 5-13 is 2 μA at room temperature (25°C) and increases by a factor of 2 for each temperature increase of 10°C. The bias $V_{BB} = 5$ V. Find the maximum allowable value for R_B if the transistor is to remain cut off at a temperature of 75°C.

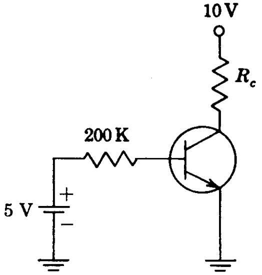
(*b*) If $V_{BB} = 1.0$ V and $R_B = 50$ K, how high may the temperature increase before the transistor comes out of cutoff?

5-10 From the characteristic curves for the type 2N404 transistor given in Fig. 5-14, find the voltages V_{BE}, V_{CE}, and V_{BC} for the circuit shown.



Prob. 5-10

5-11 A silicon transistor with $V_{BE,\text{sat}} = 0.8$ V, $\beta = h_{FE} = 100$, $V_{CE,\text{sat}} = 0.2$ V is used in the circuit shown. Find the minimum value of R_c for which the transistor remains in saturation.



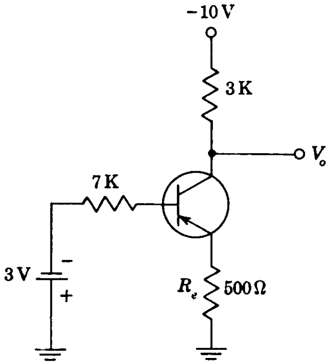
Prob. 5-11

5-12 For the circuit shown, assume $\beta = h_{FE} = 100$.

(*a*) Find if the silicon transistor is in cutoff, saturation, or in the active region.

(*b*) Find V_o.

(*c*) Find the minimum value for the emitter resistor R_e for which the transistor operates in the active region.



Prob. 5-12

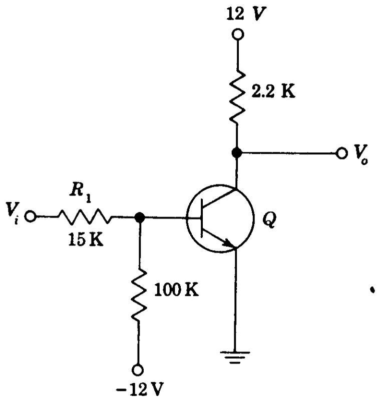
5-13 If the silicon transistor used in the circuit shown has a minimum value of $\beta = h_{FE}$ of 30 and if $I_{CBO} = 10$ nA at 25°C:

(*a*) Find V_o for $V_i = 12$ V and show that Q is in saturation.

(*b*) Find the minimum value of R_1 for which the transistor in part *a* is in the active region.

(*c*) If $R_1 = 15$ K and $V_i = 1$ V, find V_o and show that Q is at cutoff.

(*d*) Find the maximum temperature at which the transistor in part *c* remains at cutoff.

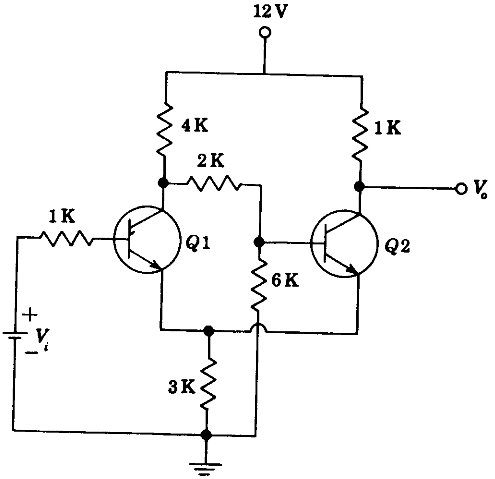


Prob. 5-13

5-14 Silicon transistors with $h_{FE} = 100$ are used in the circuit shown. Neglect the reverse saturation current.

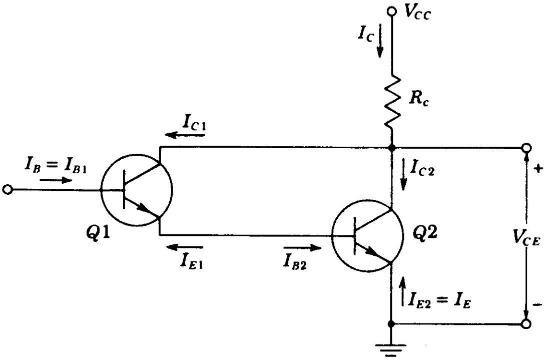
(*a*) Find V_o when $V_i = 0$ V. Assume $Q1$ is OFF and justify the assumption.

(*b*) Find V_o when $V_i = 6$ V. Assume $Q2$ is OFF and justify this assumption.



Prob. 5-14

5-15 For the circuit shown, $\alpha_1 = 0.98$, $\alpha_2 = 0.96$, $V_{CC} = 24$ V, $R_c = 120\ \Omega$, and $I_E = -100$ mA. Neglecting the reverse saturation currents, determine (*a*) the currents I_{C1}, I_{B1}, I_{E1}, I_{C2}, I_{B2}, and I_C; (*b*) V_{CE}; (*c*) I_C/I_B, I_C/I_E.



Prob. 5-15

5-16 Derive from Eqs. (5-24) and (5-25) the explicit expressions for I_C and I_E in terms of V_C and V_E.

5-17 (*a*) Derive Eqs. (5-29) and (5-30).
(*b*) Derive Eq. (5-31).

5-18 Draw the Ebers-Moll model for an *n-p-n* transistor.

5-19 (*a*) Show that the exact expression for the CE output characteristics of a *p-n-p* transistor is

$$V_{CE} = V_T \ln \frac{\alpha_I}{\alpha_N} + V_T \ln \frac{I_{CO} + \alpha_N I_B - I_C(1 - \alpha_N)}{I_{EO} + I_B + I_C(1 - \alpha_I)}$$

(*b*) Show that this reduces to Eq. (5-31) if $I_B \gg I_{EO}$ and $I_B \gg I_{CO}/\alpha_N$.

5-20 (*a*) A transistor is operating in the cutoff region with both the emitter and collector junctions reverse-biased by at least a few tenths of a volt. Prove that the currents are given by

$$I_E = \frac{I_{EO}(1 - \alpha_N)}{1 - \alpha_N\alpha_I}$$

$$I_C = \frac{I_{CO}(1 - \alpha_I)}{1 - \alpha_N\alpha_I}$$

(*b*) Prove that the emitter-junction voltage required just to produce cutoff ($I_E = 0$ and the collector back-biased) is

$$V_E = V_T \ln (1 - \alpha_N)$$

5-21 (*a*) Find the collector current for a transistor when both emitter and collector junctions are reverse-biased. Assume $I_{CO} = 5\ \mu$A, $I_{EO} = 3.57\ \mu$A, and $\alpha_N = 0.98$.
(*b*) Find the emitter current I_E under the same conditions as in part *a*.

5-22 Show that the emitter volt-ampere characteristic of a transistor in the active region is given by

$$I_E \approx I_S \epsilon^{V_E/V_T}$$

where $I_S = -I_{EO}/(1 - \alpha_N \alpha_I)$. Note that this characteristic is that of a *p-n* junction diode.

5-23 (*a*) Given an *n-p-n* transistor for which (at room temperature) $\alpha_N = 0.98$, $I_{CO} = 2\ \mu A$, and $I_{EO} = 1.6\ \mu A$. A common-emitter connection is used, and $V_{CC} = 12$ V and $R_L = 4.0$ K. What is the minimum base current required in order that the transistor enter its saturation region?

(*b*) Under the conditions in part *a*, find the voltages across each junction between each pair of terminals if the base-spreading resistance $r_{bb'}$ is neglected.

(*c*) Repeat part *b* if the base current is 200 μA.

(*d*) How are the above results modified if $r_{bb'} = 250\ \Omega$?

5-24 Plot the emitter current vs. emitter-to-base voltage for a transistor for which $\alpha_N = 0.98$, $I_{CO} = 2\ \mu A$, and $I_{EO} = 1.6\ \mu A$ if (*a*) $V_C = 0$, (*b*) V_C is back-biased by more than a few tenths of a volt. Neglect the base-spreading resistance.

5-25 Plot carefully to scale the common-emitter characteristic I_C/I_B versus V_{CE} for a transistor with $\alpha_N = 0.90 = \alpha_I$.

5-26 Show that

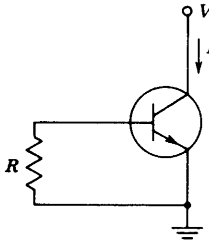
$$I_{CES} = \frac{I_{CO}}{1 - \alpha_N \alpha_I} \qquad I_{CEO} = \frac{I_{CO}}{1 - \alpha_N}$$

5-27 A common method of calculating α_N and α_I is by measurement of I_{CO}, I_{CEO}, and I_{CES}. Show that

$$(a)\ \alpha_N = \frac{I_{CEO} - I_{CO}}{I_{CEO}} \qquad (b)\ \alpha_I = \frac{1 - I_{CO}/I_{CES}}{1 - I_{CO}/I_{CEO}}$$

5-28 The collector leakage current is measured as shown in the figure, with the emitter grounded and a resistor R connected between base and ground. If this current is designated as I_{CER}, show that

$$I_{CER} = \frac{I_{CO}(1 + I_{EO}R/V_T)}{1 - \alpha_N \alpha_I + (I_{EO}R/V_T)(1 - \alpha_N)}$$



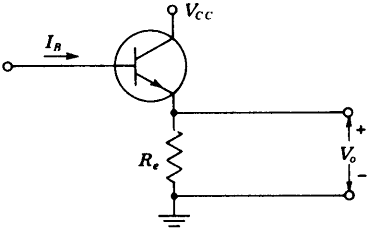
Prob. 5-28

5-29 For the circuit shown, verify that $V_o = V_{CC}$ when

$$I_B = \frac{V_{CC}}{R_e}\left(1 + \frac{\alpha_N}{\alpha_I}\frac{1-\alpha_I}{1-\alpha_N}\right) = \frac{V_{CC}}{R_e}\left(1 + \frac{\beta_N}{\beta_I}\right)$$

Under these conditions the base current exceeds the emitter current.

Prob. 5-29

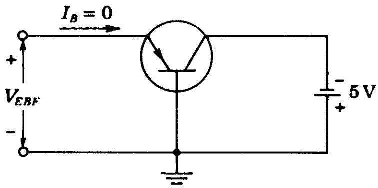


5-30 For the circuit shown, prove that the floating emitter-to-base voltage is given by

$$V_{EBF} = V_T \ln(1 - \alpha_N)$$

Neglect $r_{bb'}$.

Prob. 5-30

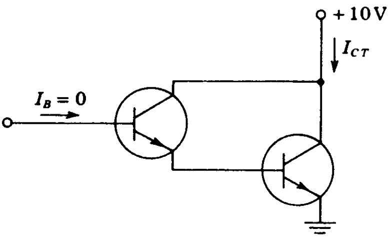


5-31 For the "floating-base" connection shown, prove that

$$I_{CT} = \frac{2 - \alpha_N}{(1-\alpha_N)^2} I_{CO}$$

Assume that the transistors are identical.

Prob. 5-31



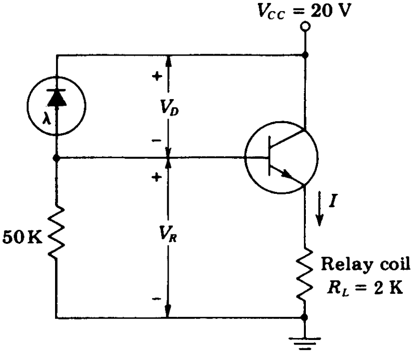
5-32 (*a*) Show that if the collector junction is reverse-biased with $|V_{CB}| \gg V_T$, the voltage V_{BE} is related to the base current by

$$V_{BE} = I_B\left(r_{bb'} + \frac{R_E}{1-\alpha_N}\right) + \left\{\frac{I_{CO}R_E}{1-\alpha_N} + V_T \ln\left[1 + \frac{I_B(1-\alpha_N\alpha_I)}{I_{EO}(1-\alpha_N)} + \frac{\alpha_N(1-\alpha_I)}{\alpha_I(1-\alpha_N)}\right]\right\}$$

where $r_{bb'}$ is the base-spreading resistance, and R_E is the emitter-body resistance. (*b*) Show that $V_{BE} = I_B(r_{bb'} + R_E) + V_T(1 + I_B/I_E)$ if the collector is open-circuited.

5-33 A transistor is operated at a forward emitter current of 2 mA and with the collector open-circuited. Find (*a*) the junction voltages V_C and V_E, (*b*) the collector-to-emitter voltage V_{CE}. Assume $I_{CO} = 2\ \mu\text{A}$, $I_{EO} = 1.6\ \mu\text{A}$, $\alpha_N = 0.98$. Is the transistor operating in saturation, at cutoff, or in the active region?

5-34 Photodiode 1N77 (Fig. 3-21) is used in the circuit shown. R_L represents the coil resistance of a relay for which the current required to close the relay is 6 mA. The transistor used is silicon with $V_{BE} = 0.7$ V and $h_{FE} = 100$.



Prob. 5-34

(*a*) Find the voltage V_D at which switching of the relay occurs.
(*b*) Find the minimum illumination required to close the relay.
(*c*) If the relay coil is placed directly in series with the phototransistor of Fig. 5-25 across 20 V, find the illumination intensity required to close the relay.

CHAPTER 6

6-1 Convert the following decimal numbers to binary form: (*a*) 671, (*b*) 325, (*c*) 152.

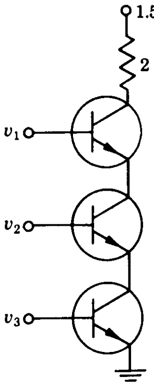
6-2 The parameters in the diode OR circuit of Fig. 6-3 are $V(0) = +12$ V, $V(1) = -2$ V, $R_s = 600\ \Omega$, $R = 10$ K, $R_f = 0$, $R_r = \infty$, and $V_\gamma = 0.6$ V. Calculate the output levels if one input is excited and if (*a*) $V_R = +12$ V, (*b*) $V_R = +10$ V, (*c*) $V_R = +14$ V, and (*d*) $V_R = 0$ V. For which of these cases is the OR function

6-48 For an RTL IC positive NOR gate prove that the maximum fan-out can be approximated by the formula

$$N_{\max} = h_{FE,\min} - h_{FE,\min}\frac{0.6}{V_{CC}} - \frac{R_b}{R_c}$$

6-49 The inputs of the RTL IC positive NOR gate shown in Fig. 6-29 are obtained from the outputs of similar gates and the outputs drive similar gates. If the supply voltage of the system is 5 V and the temperature range for proper operation of the gate is -50 to 150°C, calculate the maximum permissible values of the resistances. Assume $h_{FE} = 30$ at -50°C, $I_{CBO} = 10$ nA at 25°C, and the desired fan-out is 10.

6-50 Verify that the DCTL circuit shown with the fan-in transistors in series satisfies the NAND operation. Assume that for the silicon transistors, $V_{CE,\text{sat}} = 0.2$ V and $V_{BE,\text{sat}} = 0.8$ V. Calculate the collector currents in each transistor when all inputs are high. The input to each base is taken from the output of a similar gate.



Prob. 6-50

CHAPTER 7

7-1 (*a*) Verify that Eq. (7-3) meets the stated boundary conditions.
(*b*) Verify that Eq. (7-5) satisfies the diffusion equation (7-2) and that it meets the stated boundary conditions.

7-2 A silicon wafer is uniformly doped with phosphorus to a concentration of 10^{15} cm^{-3}. Refer to Table 2-1 on page 29. At room temperature (300°K) find
(*a*) The percentage of phosphorus by weight in the wafer.
(*b*) The conductivity and resistivity.
(*c*) The concentration of boron, which, if added to the phosphorus-doped wafer, would halve the conductivity.

7-3 (*a*) Using the data of Fig. 7-8, calculate the percent maximum concentration of arsenic (atoms per cubic centimeter) that can be achieved in solid silicon. The concentration of pure silicon may be calculated from the data in Table 2-1 on page 29.

(*b*) Repeat part *a* for gold.

7-4 (*a*) How long would it take for a fixed amount of phosphorus distributed over one surface of a 25-μm-thick silicon wafer to become substantially uniformly distributed throughout the wafer at 1300°C? Consider that the concentration is sufficiently uniform if it does not differ by more than 10 percent from that at the surface.

(*b*) Repeat part *a* for gold, given that the diffusion coefficient of gold in silicon is 1.5×10^{-6} cm^2/s at 1300°C.

7-5 Show that the junction depth x_j resulting from a Gaussian impurity diffusion into an oppositely doped material of background concentration N_{BC} is given by

$$x_j = \left(2Dt \ln \frac{Q^2}{N_{BC}^2 \pi Dt}\right)^{\frac{1}{2}}$$

7-6 A uniformly doped *n*-type silicon substrate of 0.1 Ω-cm resistivity is to be subjected to a boron diffusion with constant surface concentration of 4.8×10^{18} cm^{-3}. The desired junction depth is 2.7 μm.

(*a*) Calculate the impurity concentration for the boron diffusion as a function of distance from the surface.

(*b*) How long will it take if the temperature at which this diffusion is conducted is 1100°C?

(*c*) An *n-p-n* transistor is to be completed by diffusing phosphorus at a surface concentration of 10^{21} cm^{-3}. If the new junction is to be at a depth of 2 μm, calculate the concentration for the phosphorus diffusion as a function of distance from the surface.

(*d*) Plot the impurity concentrations (log scale) vs. distance (linear scale) for parts *a* and *c*, assuming that the boron stays put during the phosphorus diffusion. Indicate emitter, base, and collector on your plot.

(*e*) If the phosphorus diffusion takes 30 min, at what temperature is the apparatus operated?

7-7 List in order the steps required in fabricating a monolithic silicon integrated transistor by the epitaxial-diffused method. Sketch the cross section after each oxide growth. Label materials clearly. No buried layer is required.

7-8 Sketch *to scale* the cross section of a monolithic transistor fabricated on a 5-mil-thick silicon substrate. HINT: Refer to Sec. 7-1 and Figs. 7-12 and 7-13 for typical dimensions.

7-9 Sketch the five basic diode connections (in circuit form) for the monolithic integrated circuits. Which will have the lowest forward voltage drop? Highest breakdown voltage?

7-10 If the base sheet resistance can be held to within ± 10 percent and resistor line widths can be held to ± 0.1 mil, plot approximate tolerance of a diffused resistor as a function of line width w in mils over the range $0.5 \leq w \leq 5.0$. (Neglect contact-area and contact-placement errors.)

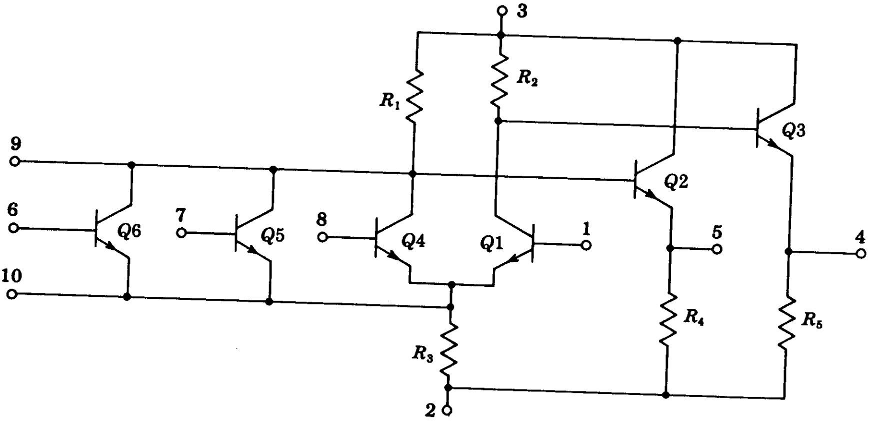
7-11 A 1-mil-thick silicon wafer has been doped uniformly with phosphorus to a concentration of 10^{16} cm^{-3}, plus boron to a concentration of 2×10^{15} cm^{-3}. Find its sheet resistance.

7-12 (*a*) Calculate the resistance of a diffused crossover 4 mils long, 1 mil wide, and 2 μm thick, given that its sheet resistance is 2.2 Ω/square.

(*b*) Repeat part *a* for an aluminum metalizing layer 0.5 μm thick of resistivity 2.8×10^{-6} Ω-cm. Note the advantage of avoiding diffused crossovers.

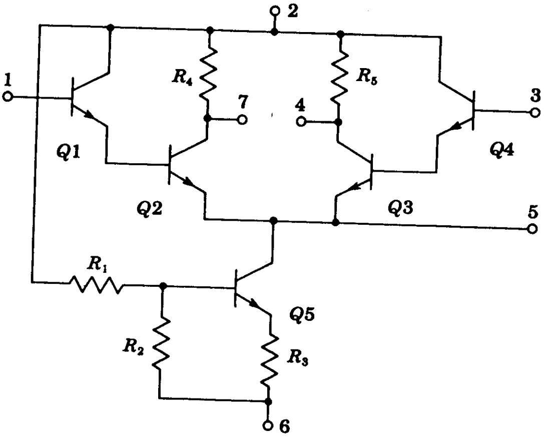
7-13 (*a*) What is the minimum number of isolation regions required to realize in monolithic form the logic gate shown?

(*b*) Draw a monolithic layout of the gate in the fashion of Fig. 7-25*b*.


3
R_2
R_1
$Q3$
9
$Q2$
6
$Q6$
7
$Q5$
8
$Q4$
$Q1$
1
5
4
10
R_4
R_5
R_3
2

Prob. 7-13

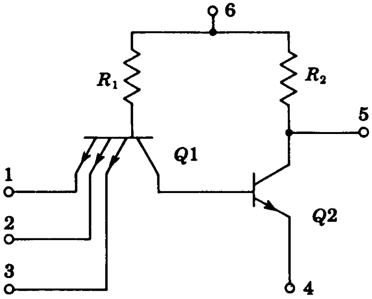
7-14 Repeat Prob. 7-13 for the difference amplifier shown.


2
1
R_4
R_5
7
4
3
$Q1$
$Q4$
$Q2$
$Q3$
5
R_1
$Q5$
R_2
R_3
6

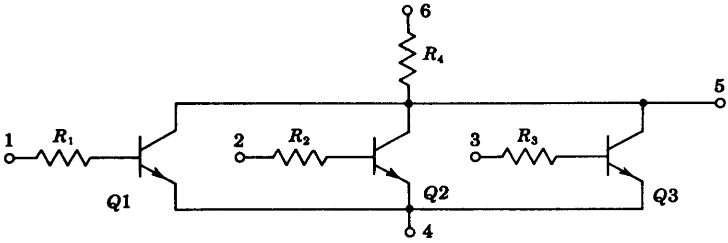
Prob. 7-14

7-15 For the circuit shown, find (*a*) the *minimum* number, (*b*) the *maximum* number, of isolation regions.

Prob. 7-15



7-16 For the circuit shown, (*a*) find the minimum number of isolation regions, and (*b*) draw a monolithic layout in the fashion of Fig. 7-26, given that (i) $Q1$, $Q2$, and $Q3$ should be single-base-stripe, 1- by 2-mil emitter, transistors, (ii) $R_1 = R_2 = R_3 = 400\ \Omega$, $R_4 = 600\ \Omega$. Use 1-mil-wide resistors.

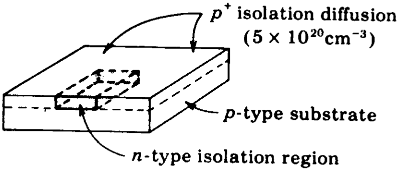


Prob. 7-16

7-17 An integrated junction capacitor has an area of 1,000 mils^2 and is operated at a reverse barrier potential of 1 V. The acceptor concentration of 10^{15} atoms/cm^3 is much smaller than the donor concentration. Calculate the capacitance.

7-18 A thin-film capacitor has a capacitance of 0.4 pF/mil^2. The relative dielectric constant of silicon dioxide is 3.5. What is the thickness of the SiO_2 layer in angstroms?

7-19 The *n*-type epitaxial isolation region shown is 8 mils long, 6 mils wide, and 1 mil thick and has a resistivity of 0.1 Ω-cm. The resistivity of the *p*-type substrate is 10 Ω-cm. Find the parasitic capacitance between the isolation region and the substrate under 5-V reverse bias. Assume that the sidewalls contribute 0.1 pF/mil^2.


p^+ isolation diffusion
(5×10^{20}cm^{-3})
p-type substrate
n-type isolation region

Prob. 7-19

NOTE: *In the problems that follow, indicate your answer by giving the letter of the statement you consider correct.*

7-20 The typical number of diffusions used in making epitaxial-diffused silicon integrated circuits is (*a*) 1, (*b*) 2, (*c*) 3, (*d*) 4, (*e*) 5.

7-21 The "buried layer" in an integrated transistor is (*a*) p^+ doped, (*b*) located in the base region, (*c*) n^+ doped, (*d*) used to reduce the parasitic capacitance.

7-22 Epitaxial growth is used in integrated circuits (ICs)

(*a*) To grow selectively single-crystal p-doped silicon of one resistivity on a p-type substrate of a different resistivity.

(*b*) To grow single-crystal n-doped silicon on a single-crystal p-type substrate.

(*c*) Because it yields back-to-back isolating p-n junctions.

(*d*) Because it produces low parasitic capacitance.

7-23 Silicon dioxide (SiO_2) is used in ICs

(*a*) Because it facilitates the penetration of diffusants.

(*b*) Because of its high heat conduction.

(*c*) To control the location of diffusion and to protect and insulate the silicon surface.

(*d*) To control the concentration of diffusants.

7-24 The p-type substrate in a monolithic circuit should be connected to

(*a*) The most positive voltage available in the circuit.

(*b*) The most negative voltage available in the circuit.

(*c*) Any dc ground point.

(*d*) Nowhere, i.e., be left floating.

7-25 Monolithic integrated circuit systems offer greater reliability than discrete-component systems because

(*a*) There are fewer interconnections.

(*b*) High-temperature metalizing is used.

(*c*) Electric voltages are low.

(*d*) Electric elements are closely matched.

7-26 The collector-substrate junction in the epitaxial collector structure is, approximately,

(*a*) A step-graded junction.

(*b*) A linearly graded junction.

(*c*) An exponential junction.

(*d*) None of the above.

7-27 The sheet resistance of a semiconductor is

(*a*) An undesirable parasitic element.

(*b*) An important characteristic of a diffused region, especially when used to form diffused resistors.

(*c*) A characteristic whose value determines the required area for a given value of integrated capacitance.
(*d*) A parameter whose value is important in a thin-film resistance.

7-28 Isolation in ICs is required.
(*a*) To make it simpler to test circuits.
(*b*) To protect the components from mechanical damage.
(*c*) To protect the transistor from possible "thermal runaway."
(*d*) To minimize electrical interaction between circuit components.

7-29 Almost all resistors are made in a monolithic IC
(*a*) During the emitter diffusion.
(*b*) While growing the epitaxial layer.
(*c*) During the base diffusion.
(*d*) During the collector diffusion.

7-30 Increasing the yield of an integrated circuit
(*a*) Reduces individual circuit cost.
(*b*) Increases the cost of each good circuit.
(*c*) Results in a lower number of good chips per wafer.
(*d*) Means that more transistors can be fabricated on the same size wafer.

7-31 In a monolithic-type IC
(*a*) All isolation problems are eliminated.
(*b*) Resistors and capacitors of any value may be made.
(*c*) All components are fabricated into a single crystal of silicon.
(*d*) Each transistor is diffused into a separate isolation region.

7-32 The main purpose of the metalization process is
(*a*) To interconnect the various circuit elements
(*b*) To protect the chip from oxidation.
(*c*) To act as a heat sink.
(*d*) To supply a bonding surface for mounting the chip.

CHAPTER 8

NOTE: *Unless otherwise specified, all transistors in these problems are identical, and the numerical values of their h parameters are given in Table* 8-2. *Also assume that all capacitances are arbitrarily large.*

8-1 (*a*) Using Fig. 8-6*b* write the input and output equations.
(*b*) Draw the hybrid model for a CB transistor and write the input and output equations.

8-2 (*a*) Describe how to obtain h_{ie} from the CE input characteristics.
(*b*) Repeat part *a* for h_{re}. Explain why this procedure, although correct in principle, is inaccurate in practice.

8-3 The transistor whose input characteristics are shown in Fig. 8-2 is biased at $V_{CE} = -8$ V and $I_B = -300\ \mu$A.
(*a*) Compute graphically h_{fe} and h_{oe} at the quiescent point specified above.
(*b*) Using the *h* parameters computed in part *a*, calculate h_{fb} and h_{ob}.

8-4 Justify the statement in the footnote to Table 8-3. HINT: Draw a CE transistor circuit with a signal voltage V_s between base and ground. Now interchange *B* and *E* and observe the resulting configuration.

(*b*) For the circuit of part *a* determine the variation of I_C in the temperature range -65 for 175°C when the silicon transistor of Table 9-1 is used.

9-29 Prove Eq. (9-38).

9-30 (*a*) The circuit of Prob. 9-19 is modified by the addition of a thermistor as in Fig. 9-12. Find R_T, I_T, and V_{CE} for the modified circuit if $I_C = 1.5$ mA and $V_{CC} = 27.5$ V.

(*b*) It is desired that as the temperature changes from 25 to 175°, the variation of I_C be $+0.4$ mA. Calculate the temperature coefficient of the thermistor.

9-31 (*a*) Calculate the thermal resistance for the 2N338 transistor for which the manufacturer specifies $P_{C,\max} = 125$ mW at 25°C free-air temperature and maximum junction temperature $T_j = 150$°C.

(*b*) What is the junction temperature if the collector dissipation is 75 mW?

9-32 Show that the load line tangent to the constant-power-dissipation hyperbola of Fig. 9-14 is bisected by the tangency point, that is, $AC = BC$.

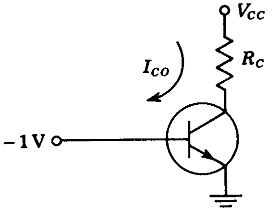
9-33 The transistor used in the circuit is at cutoff.

(*a*) Show that runaway will occur for values of I_{CO} in the range

$$\frac{V_{CC} - \sqrt{V_{CC}^2 - 8R_c/0.07\Theta}}{4R_c} \leq I_{CO} \leq \frac{V_{CC} + \sqrt{V_{CC}^2 - 8R_c/0.07\Theta}}{4R_c}$$

(*b*) Show that if runaway is not destructive, the collector current I_{CO} after runaway can never exceed $I_{CO} = V_{CC}/2R_c$.

Prob. 9-33



9-34 A germanium transistor with $\Theta = 250$°C/W, $I_{CO} = 10\ \mu$A at 25°C, $R_c = 1$ K, and $V_{CC} = 30$ V is used in the circuit of Prob. 9-33.

(*a*) Find I_{CO} at the point of runaway.

(*b*) Find the ambient temperature at which runaway will occur.

CHAPTER 10

10-1 The drain resistance R_d of an *n*-channel FET with the source grounded is 2 K. The FET is operating at a quiescent point $V_{DS} = 10$ V, and $I_{DS} = 3$ mA, and its characteristics are given in Fig. 10-3.

(*a*) To what value must the gate voltage be changed if the drain current is to change to 5 mA?

(*b*) To what value must the drain voltage be changed if the drain current is to be brought back to its previous value? The gate voltage is maintained constant at the value found in part *a*.

10-2 For a *p*-channel silicon FET with $a = 2 \times 10^{-4}$ cm and channel resistivity $\rho = 10\ \Omega$-cm

(*a*) Find the pinch-off voltage.

(*b*) Repeat (*a*) for a *p*-channel germanium FET with $\rho = 2\ \Omega$-cm.

10-3 (*a*) Plot the transfer characteristic curve of an FET as given by Eq. (10-8), with $I_{DSS} = 10$ mA and $V_P = -4$ V.

(*b*) The magnitude of the slope of this curve at $V_{GS} = 0$ is g_{mo} and is given by Eq. (10-17). If the slope is extended as a tangent, show that it intersects the V_{GS} axis at the point $V_{GS} = V_P/2$.

10-4 (*a*) Show that the transconductance g_m of a JFET is related to the drain current I_{DS} by

$$g_m = \frac{2}{|V_P|}\sqrt{I_{DSS}I_{DS}}$$

(*b*) If $V_P = -4$ V and $I_{DSS} = 4$ mA, plot g_m versus I_{DS}.

10-5 Show that for small values of V_{GS} compared with V_P, the drain current is given approximately by $I_D \approx I_{DSS} + g_{mo}V_{GS}$.

10-6 (*a*) For the FET whose characteristics are plotted in Fig. 10-3, determine r_d and g_m graphically at the quiescent point $V_{DS} = 10$ V and $V_{GS} = -1.5$ V. Also evaluate μ.

(*b*) Determine $r_{d,\text{ON}}$ for $V_{GS} = 0$.

10-7 (*a*) Verify Eq. (10-15).

(*b*) Starting with the definitions of g_m and r_d, show that if two identical FETs are connected in parallel, g_m is doubled and r_d is halved. Since $\mu = r_d g_m$, then μ remains unchanged.

(*c*) If the two FETs are not identical, show that

$$\frac{1}{r_d} = \frac{1}{r_{d1}} + \frac{1}{r_{d2}}$$

and that

$$\mu = \frac{\mu_1 r_{d2} + \mu_2 r_{d1}}{r_{d1} + r_{d2}}$$

10-8 Given the transfer characteristic of an FET, explain clearly how to determine g_m at a specified quiescent point.

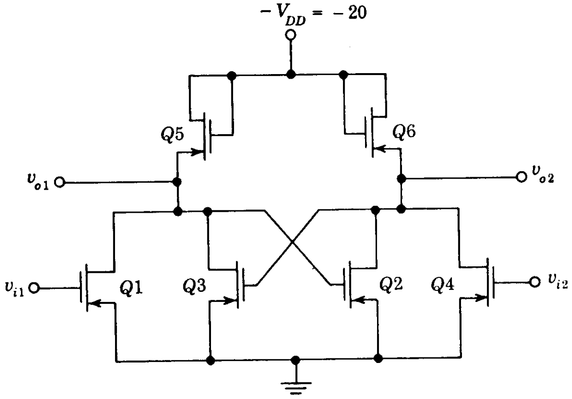
10-9 (*a*) Using Eq. (10-18), prove that the load conductance of a MOSFET is given by

$$g_L = g_{do}\left(1 - \frac{V_{DS}}{V_T}\right) \qquad \text{where } g_{do} = \frac{2I_{DSS}}{|V_T|}$$

(*b*) Prove that $g_L = g_m$ for a JFET. HINT: Use Eq. (10-8).

10-10 Draw the circuit of a MOSFET negative AND gate and explain its operation.

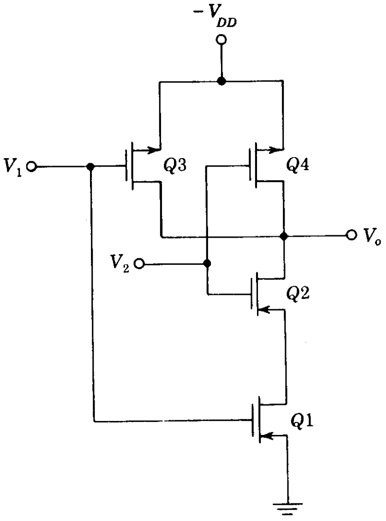
10-11 Consider the FLIP-FLOP circuit shown. Assume $V_T = V_{ON} = 0$ and $|V_{DD}| \gg V_T$.
(*a*) Assume $v_{i1} = v_{i2} = 0$. Verify that the circuit has two possible stable states; either $v_{o1} = 0$ and $v_{o2} = -V_{DD}$ or $v_{o1} = -V_{DD}$ and $v_{o2} = 0$.
(*b*) Show that the state of the FLIP-FLOP may be changed by momentarily allowing one of the inputs to go to $-V_{DD}$; in other words by applying a negative input pulse.



Prob. 10-11

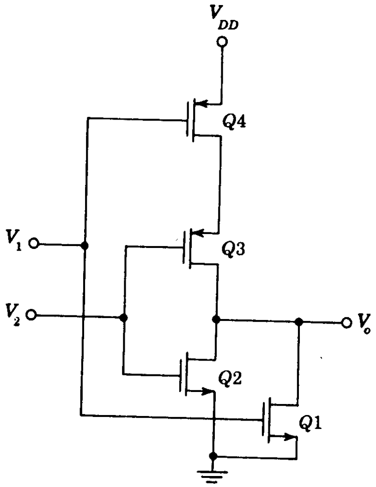
10-12 Draw a CMOS inverter using positive logic.

10-13 (*a*) The complementary MOS negative NAND gate is indicated. Explain its operation.
(*b*) Draw the corresponding positive NAND gate.



Prob. 10-13

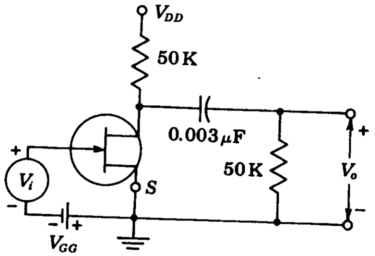
10-14 The circuit of a CMOS positive NOR gate is indicated. Explain its operation.



Prob. 10-14

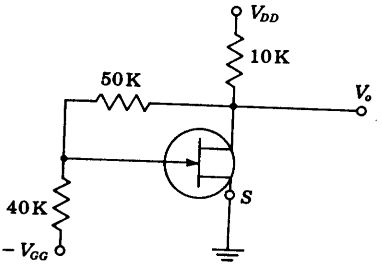
10-15 Draw a MOSFET circuit satisfying the logic equation. $Y = \overline{A + BC}$, where Y is the output corresponding to the three inputs A, B, and C.

10-16 (*a*) Calculate the voltage gain $A_V = V_o/V_i$ at 1 kHz for the circuit shown. The FET parameters are $g_m = 2$ mA/V and $r_d = 10$ K. Neglect capacitances. (*b*) Repeat part *a* if the capacitance 0.003 μF is taken under consideration.



Prob. 10-16

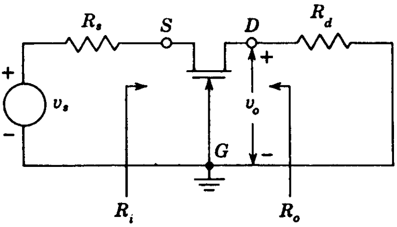
10-17 If an input signal V_i is impressed between gate and ground, find the amplification $A_V = V_o/V_i$. Apply Miller's theorem to the 50-K resistor. The FET parameters are $\mu = 30$ and $r_d = 5$ K. Neglect capacitances.



Prob. 10-17

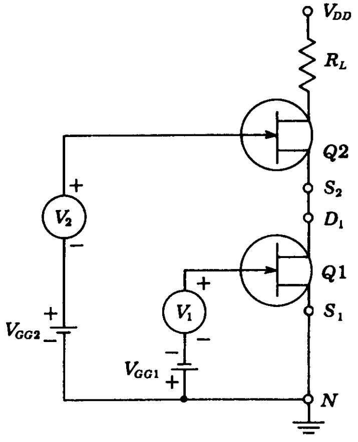
10-18 If in Prob. 10-17 the signal V_i is impressed in series with the 40-K resistor (instead of from gate to ground), find $A_V = V_o/V_i$.

10-19 The circuit shown is called common-gate amplifier. For this circuit find (*a*) the voltage gain, (*b*) the input impedance, (*c*) the output impedance. Power supplies are omitted for simplicity. Neglect capacitances.

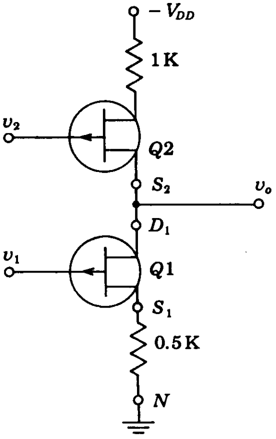


Prob. 10-19

10-20 Find an expression for the signal voltage across R_L. The two FETs are identical, with parameters μ, r_d, and g_m. HINT: Use the equivalent circuits in Fig. 10-22 at S_2 and D_1.



Prob. 10-20



Prob. 10-21

10-21 Each FET shown has the parameters $r_d = 10$ K and $g_m = 2$ mA/V. Using the equivalent circuits in Fig. 10-22 at S_2 and D_1, find the gain (*a*) v_o/v_1 if $v_2 = 0$, (*b*) v_o/v_2 if $v_1 = 0$.

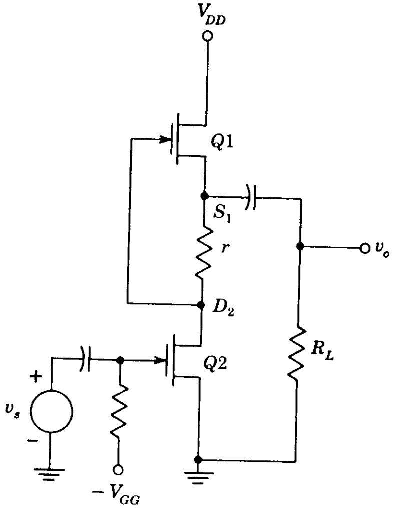
10-22 (*a*) Prove that the magnitude of the signal current is the same in both FETs provided that

$$r = \frac{1}{g_m} + \frac{2R_L}{\mu}$$

Neglect the reactance of the capacitors.

(*b*) If r is chosen as in part *a*, prove that the voltage gain is given by

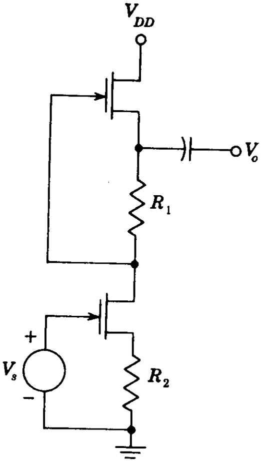
$$A = \frac{-\mu^2}{\mu + 1} \frac{R_L}{R_L + r_d/2}$$



Prob. 10-22

10-23 (*a*) If $R_1 = R_2 = R$ and the two FETs have identical parameters, verify that the voltage amplification is $V_o/V_s = -\mu/2$ and the output impedance is $\frac{1}{2}[r_d + (\mu + 1)R]$.

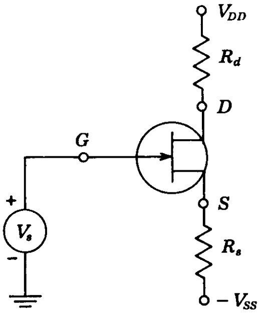
(*b*) Given $r_d = 62$ K, $\mu = 10$, $R_1 = 2$ K, and $R_2 = 1$ K. Find the voltage gain and the output impedance.



Prob. 10-23

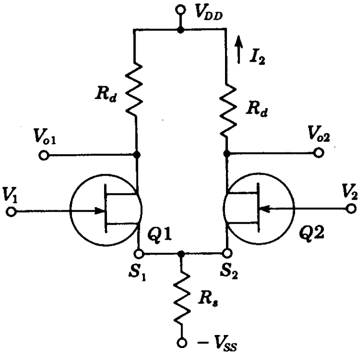
10-24 (*a*) If in the amplifier stage shown the positive supply voltage V_{DD} changes by $\Delta V_{DD} = v_a$, how much does the drain-to-ground voltage change?
(*b*) How much does the source-to-ground voltage change under the conditions in part *a*?
(*c*) Repeat parts *a* and *b* if V_{DD} is constant but V_{SS} changes by $\Delta V_{SS} = v_s$.

Prob. 10-24



10-25 If in the circuit shown $V_2 = 0$, then this circuit becomes a source-coupled phase inverter, since $V_{o1} = -V_{o2}$. Solve for the current I_2 by drawing the equivalent circuit, looking into the source of $Q1$ (Fig. 10-22). Then replace $Q2$ by the equivalent circuit, looking into its drain. The source resistance R_s may be taken as arbitrarily large.

Prob. 10-25

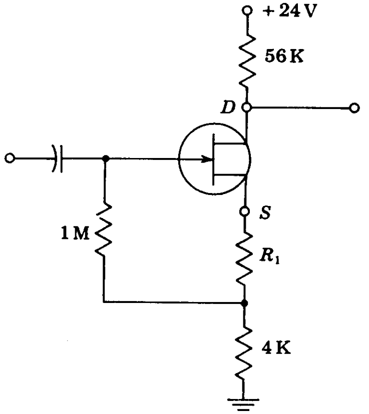


10-26 In the circuit of Prob. 10-25, assume that $V_2 = 0$, $R_d = r_d = 10$ K, $R_s = 1$ K, and $\mu = 19$. If the output is taken from the drain of $Q2$, find (*a*) the voltage gain, (*b*) the output impedance. HINT: Use the equivalent circuits in Fig. 10-22.

10-27 In the circuit of Prob. 10-25, $V_2 \neq V_1$, $R_d = 30$ K, $R_s = 2$ K, $\mu = 19$, and $r_d = 10$ K. Find (*a*) the voltage gains A_1 and A_2 defined by $V_{o2} = A_1V_1 + A_2V_2$. HINT: Use the equivalent circuits in Fig. 10-22. (*b*) If R_s is arbitrarily large, show that $A_2 = -A_1$. Note that the circuit now behaves as a difference amplifier.

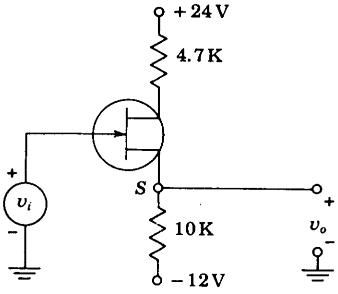
10-28 The CS amplifier stage shown in Fig. 10-23 has the following parameters: $R_d = 12$ K, $R_g = 1$ M, $R_s = 470\ \Omega$, $V_{DD} = 30$ V, C_s is arbitrarily large, $I_{DSS} = 3$ mA, $V_P = -2.4$ V, and $r_d \gg R_d$. Determine (*a*) the gate-to-source bias voltage V_{GS}, (*b*) the drain current I_D, (*c*) the quiescent voltage V_{DS}, (*d*) the small-signal voltage gain A_V.

10-29 The amplifier stage shown uses an *n*-channel FET having $I_{DSS} = 1$ mA, $V_P = -1$ V. If the quiescent drain-to-ground voltage is 10 V, find R_1.

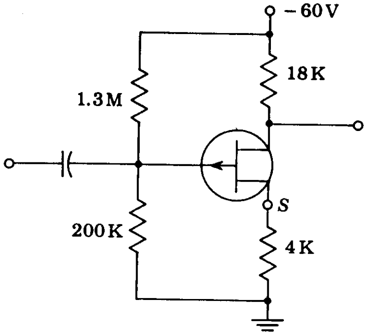


Prob. 10-29

10-30 The FET shown has the following parameters: $I_{DSS} = 5.6$ mA and $V_P = -4$ V.
(*a*) If $v_i = 0$, find v_o.
(*b*) If $v_i = 10$ V, find v_o.
(*c*) If $v_o = 0$, find v_i.
NOTE: v_i and v_o are constant voltages (and not small-signal voltages).



Prob. 10-30



Prob. 10-31

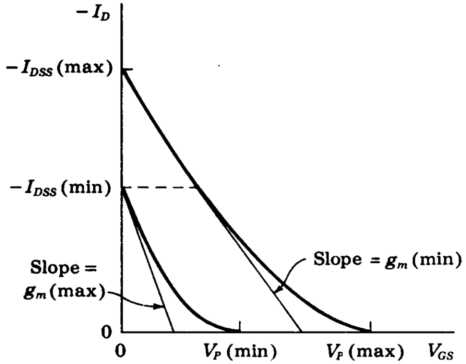
10-31 If $|I_{DSS}| = 4$ mA, $V_P = 4$ V, calculate the quiescent values of I_D, V_{GS}, and V_{DS}.

10-32 In the figure shown, two extreme transfer characteristics are indicated. The values of $V_{P,\max}$ and $V_{P,\min}$ are difficult to determine accurately. Hence these values are calculated from the experimental values of $I_{DSS,\max}$, $I_{DSS,\min}$, $g_{m,\max}$, and $g_{m,\min}$. Note that g_m is the slope of the transfer curve and that both $g_{m,\max}$ and $g_{m,\min}$ are measured at a drain current corresponding to $I_{DSS,\min}$. Verify that

(a) $V_{P,\max} = -\dfrac{2}{g_{m,\min}}(I_{DSS,\max}I_{DSS,\min})^{\frac{1}{2}}$

(b) $V_{P,\min} = -\dfrac{2I_{DSS,\min}}{g_{m,\max}}$

(c) If for a given FET, $I_{DSS,\min} = 2$ mA, $I_{DSS,\max} = 6$ mA, $g_{m,\min} = 1.5$ mA/V, and $g_{m,\max} = 3$ mA/V, evaluate $V_{P,\max}$ and $V_{P,\min}$.

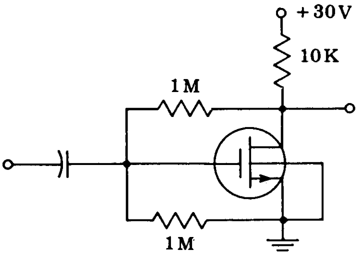


Prob. 10-32

10-33 The drain current in milliamperes of the enhancement-type MOSFET shown is given by

$$I_D = 0.2(V_{GS} - V_P)^2$$

in the region $V_{DS} \geq V_{GS} - V_P$. If $V_P = +3$ V, calculate the quiescent values I_D, V_{GS}, and V_{DS}.

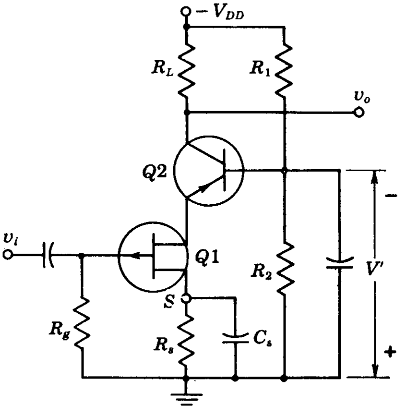


Prob. 10-33

10-34 Show that if $R_L \ll 1/h_{ob2}$, the voltage gain of the hybrid cascode amplifier stage shown is given to a very good approximation by

$$A_V = g_m h_{fb} R_L$$

where g_m is the FET transconductance.

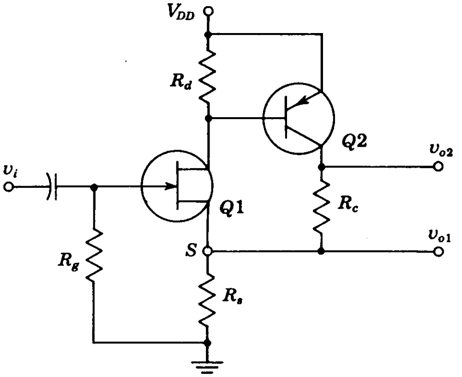


Prob. 10-34

10-35 If $h_{ie} \ll R_d$, $h_{ie} \ll r_d$, $h_{fe} \gg 1$, and $\mu \gg 1$ for the circuit, show that

$$(a)\ A_{V1} = \frac{v_{o1}}{v_i} \approx \frac{g_m h_{fe} R_s}{1 + g_m h_{fe} R_s} \qquad (b)\ A_{V2} = \frac{v_{o2}}{v_i} \approx \frac{g_m h_{fe}(R_s + R_c)}{1 + g_m h_{fe} R_s}$$

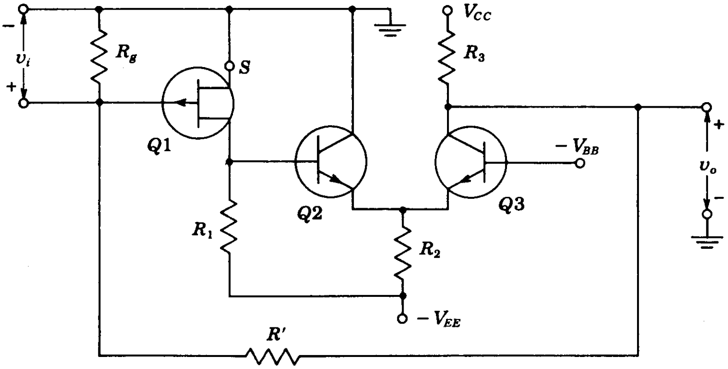
where g_m is the FET transconductance.



Prob. 10-35

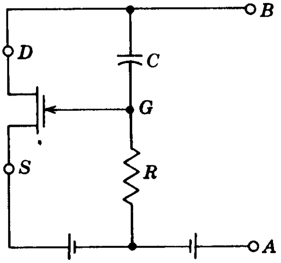
10-36 If $r_d \gg R_1$, $R_2 \gg h_{ib3}$, $1/h_{oe2} \gg h_{ib3}$, $R' \gg R_3$, and $1/h_{ob3} \gg R_3$, show that the voltage gain at low frequencies is given by

$$A_o = \frac{v_o}{v_i} = g_m(1 + h_{fe2})h_{fb3}\frac{R_1 R_3}{R_1 + h_{ie2} + h_{ib3}(1 + h_{fe2})}$$



Prob. 10-36

10-37 In the circuit shown, the FET is used as an adjustable impedance element by varying the dc bias, and thereby the g_m of the FET.
(*a*) Assume that there is a generator V between the terminals A and B. Draw the equivalent circuit. Neglect interelectrode capacitances.



Prob. 10-37

(*b*) Show that the input admittance between A and B is

$$Y_i = Y_d + (1 + g_m R) Y_{CR}$$

where Y_d is the admittance corresponding to r_d, and Y_{CR} is the admittance corresponding to R and C in series.
(*c*) If $g_m R \gg 1$, show that the effective input capacitance is

$$C_i = \frac{g_m \alpha}{\omega(1 + \alpha^2)}$$

and the effective input resistance is

$$R_i = \frac{(1 + \alpha^2) r_d}{1 + \alpha^2(1 + \mu)}$$

where $\alpha \equiv \omega C R$.

(*d*) At a given frequency, show that the maximum value of C_i (as either C or R is varied) is obtained when $\alpha = 1$, and

$$(C_i)_{\max} = \frac{g_m}{2\omega}$$

Also show that the value of R_i corresponding to this C_i is

$$(R_i)_{\max} = \frac{2r_d}{2 + \mu}$$

which, for $\mu \gg 2$, reduced to $(R_i)_{\max} = 2/g_m$.

10-38 Solve Prob. 10-37 if the capacitance C is replaced by an inductance L.

10-39 (*a*) A MOSFET connected in the CS configuration works into a 100-K resistive load. Calculate the complex voltage gain and the input admittance of the system for frequencies of 100 and 100,000 Hz. Take the interelectrode capacitances into consideration. The MOSFET parameters are $\mu = 100$, $r_d = 40$ K, $g_m = 2.5$ mA/V, $C_{gs} = 4.0$ pF, $C_{ds} = 0.6$ F, and $C_{gd} = 2.4$ pF. Compare these results with those obtained when the interelectrode capacitances are neglected.
(*b*) Calculate the input resistance and capacitance.

10-40 Calculate the input admittance of an FET at 10^3 and 10^6 Hz when the total drain circuit impedance is (*a*) a resistance of 50 K, (*b*) a capacitive reactance of 50 K at each frequency. Take the interelectrode capacitances into consideration. The FET parameters are $\mu = 20$, $r_d = 10$ K, $g_m = 2.0$ mA/V, $C_{gs} = 3.0$ pF, $C_{ds} = 1.0$ pF, and $C_{gd} = 2.0$ pF. Express the results in terms of the input resistance and capacitance.

10-41 (*a*) Starting with the circuit model of Fig. 10-31, verify Eq. (10-37) for the voltage gain of the source follower, taking interelectrode capacitances into account.
(*b*) Verify Eq. (10-39) for the input admittance.
(*c*) Verify Eq. (10-40) for the output admittance.
HINT: For part *c*, set $V_i = 0$ and impress an external voltage V_o from S to N; the current drawn from V_o divided by V_o is Y_o.

10-42 Starting with the circuit model of Fig. 10-8, show that, for the *CG* amplifier stage with $R_s = 0$ and $C_{ds} = 0$,

(*a*) $A_V = \dfrac{(g_m + g_d)R_d}{1 + R_d(g_d + j\omega C_{gd})}$ (*b*) $Y_i = g_m + g_d(1 - A_V) + j\omega C_{sg}$

(*c*) Repeat (*a*), taking the source resistance R_s into account.
(*d*) Repeat (*b*), taking the source resistance R_s into account.

10-43 (*a*) For the source follower with $g_m = 2$ mA/V, $R_s = 100$ K, $r_d = 50$ K, and with each internode capacitance 3 pF, find the frequency at which the reactive component of the output admittance equals the resistive component.
(*b*) At the frequency found in part *a* calculate the gain and compare it with the low-frequency value.

Find the value of this limiting frequency before the output signal is distorted by the slew-rate limit if (*a*) $V_m = 1$ V, (*b*) $V_m = 10$ V.

CHAPTER 16

16-1 Design the circuit of Fig. 16-1 so that the output V_o (for a sinusoidal signal) is equal in magnitude to the input V_s and leads the input by 45°.

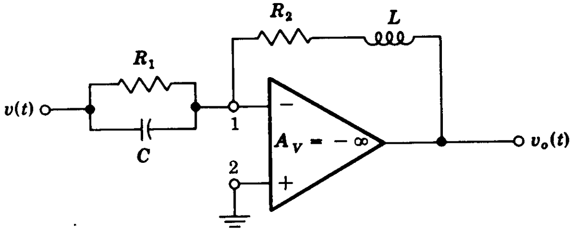
16-2 Consider the circuit of Fig. 16-1 with $A_V = -100$. If $Z = R$ and $Z' = -jX_C$ with $R = X_C$ at some specific frequency f, calculate the gain V_o/V_s as a complex number.

16-3 Given the operational amplifier circuit of Fig. 16-1 consisting of R and L in series for Z, and C for Z'. If the input voltage is a constant $v_s = V$, find the output v_o as a function of time. Assume an infinite open-loop gain.

16-4 For the given circuit, show that the output voltage is

$$-v_o = \frac{R_2}{R_1}v + \left(R_2C + \frac{L}{R_1}\right)\frac{dv}{dt} + LC\frac{d^2v}{dt^2}$$

Prob. 16-4



16-5 Consider the operational amplifier circuit of Fig. 16-1 with Z consisting of a resistor R in parallel with a capacitor C, and Z' consisting of a resistor R'. The input is a sweep voltage $v = \alpha t$. Show that the output voltage v_o is a sweep voltage that starts with an initial step. Thus prove that

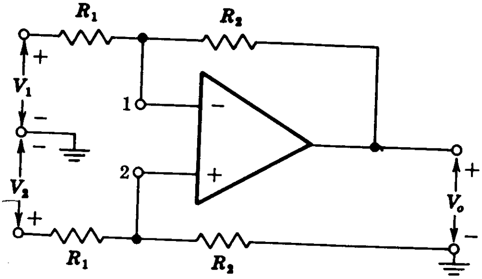
$$v_o = -\alpha R'C - \alpha\frac{R'}{R}t$$

Assume infinite open-loop gain.

16-6 Consider the operational amplifier circuit of Fig. 16-1 with Z consisting of a 100-K resistor and a series combination of a 50-K resistance with a 0.001-μF capacitance for Z'. If the capacitor is initially uncharged, and if at $t = 0$ the input voltage $v_s = 10\epsilon^{-t/\tau}$ with $\tau = 5 \times 10^{-4}$ s is applied, find $v_o(t)$.

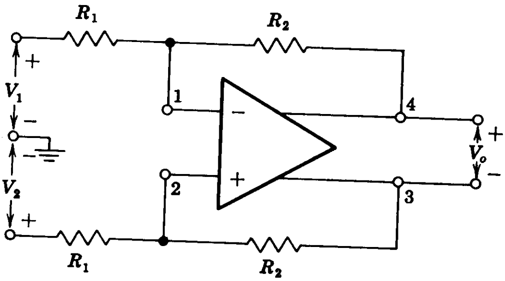
16-7 In Fig. 16-3*b* show that i_L is equal to $-v_s/R_2$ if $R_3/R_2 = R'/R_1$.

16-8 The differential input operational amplifier shown consists of a base amplifier of infinite gain. Show that $V_o = (R_2/R_1)(V_2 - V_1)$.


R_1
R_2
V_1
V_2
1
2
−
+
V_o
R_1
R_2

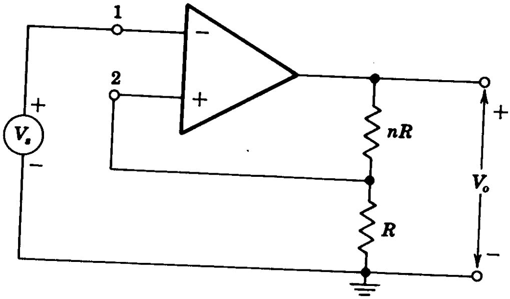
Prob. 16-8

16-9 Repeat Prob. 16-8 for the amplifier shown.


R_1
R_2
V_1
V_2
1
2
3
4
−
+
V_o
R_1
R_2

Prob. 16-9

16-10 For the base differential-input amplifier shown, assume infinite input resistance, zero output resistance, and finite differential gain $A_V = V_o/(V_1 - V_2)$.
(*a*) Obtain an expression for the gain $A_{Vf} = V_o/V_s$.
(*b*) Show that $\lim A_{Vf} = n + 1$, $A_V \to \infty$.

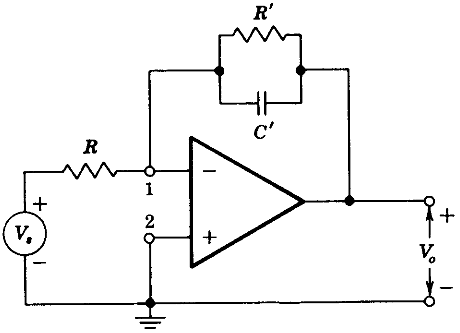


Prob. 16-10

16-11 Verify Eq. (16-6) for the bridge amplifier.

16-12 The circuit shown represents a low-pass dc-coupled amplifier. Assuming an ideal operational amplifier determine (*a*) the high-frequency 3-dB point f_H; (*b*) the low-frequency gain $A_V = V_o/V_s$.

Prob. 16-12



16-13 (*a*) The input to the operational integrator of Fig. 16-10 is a step voltage of magnitude V. Show that the output is

$$v_o = A_V V(1 - \epsilon^{-t/RC(1-A_V)})$$

(*b*) Compare this result with the output obtained if the step voltage is impressed upon a simple *RC* integrating network (without the use of an operational amplifier). Show that for large values of *RC*, both solutions represent a voltage which varies approximately linearly with time. Verify that if $-A_V \gg 1$, the slope of the ramp output is approximately the same for both circuits. Also prove that the deviation from linearity for the amplifier circuit is $1/(1 - A_V)$ times that of the simple *RC* circuit.

16-14 Derive Eq. (16-13).

16-15 (*a*) The input to an operational differentiator whose open-loop gain $A_V \equiv A$ is infinite is a ramp voltage $v = \alpha t$. Show that the output is

$$v_o = \frac{A}{1 - A} \alpha RC(1 - \epsilon^{-t(1-A)/RC})$$

(*b*) Compare this result with that obtained if the same input is impressed upon a simple *RC* differentiating network (without the use of an amplifier). Show that, approximately, the same final constant output $RC\,dv/dt$ is obtained. Also show that the operational-amplifier output reaches this correct value of the differentiated input much more quickly than does the simple *RC* circuit.

16-16 Given an operational amplifier with Z consisting of R in series with C, and Z' consisting of R' in parallel with C'. The input is a step voltage of magnitude V. (*a*) Show by qualitative argument that the output voltage must start at zero, reach a maximum, and then again fall to zero.

(*b*) Show that if $R'C' \neq RC$, the output is given by

$$v_o = \frac{R'CV}{R'C' - RC}(\epsilon^{-t/RC} - \epsilon^{-t/R'C'})$$

16-17 Sketch an operational amplifier circuit having an input v and an output which is approximately $-5v - 3dv/dt$. Assume an ideal operational amplifier.

16-18 Sketch in block-diagram form a computer, using operational amplifiers, to solve the differential equation

$$\frac{dv}{dt} + 0.5v + 0.1 \sin \omega t = 0$$

An oscillator is available which will provide a signal sin ωt. Use only resistors and capacitors.

16-19 Set up a computer in block-diagram form, using operational amplifiers, to solve the following differential equation:

$$\frac{d^3y}{dt^3} + 2\frac{d^2y}{dt^2} - 4\frac{dy}{dt} + 2y = x(t)$$

where

$$y(0) = 0 \qquad \left.\frac{dy}{dt}\right|_{t=0} = -2 \qquad \text{and} \qquad \left.\frac{d^2y}{dt^2}\right|_{t=0} = 3$$

Assume that a generator is available which will provide the signal $x(t)$.

16-20 (*a*) Verify that the damping factor of each pair of complex poles of a Butterworth low-pass filter is given by $k = \cos \theta$, where θ is defined in Fig. 16-17.
(*b*) Define a damping factor k for the single pole at $s = -1$ which is consistent with Eq. (16-23).

16-21 Verify the entries in Table 16-1 for $n = 3$ by using Fig. 16-17*b*.

16-22 Using Eq. (16-22), show that the transfer function of a second-order Butterworth low-pass filter satisfies Eq. (16-19).

16-23 Use the value of $P_2(s)$ from Table 16-1 and verify that

$$P_2(s)P_2(-s)\Big|_{s=j\omega} = 1 + \omega^4$$

16-24 Use the values of $P_3(s)$ from Table 16-1 and verify that

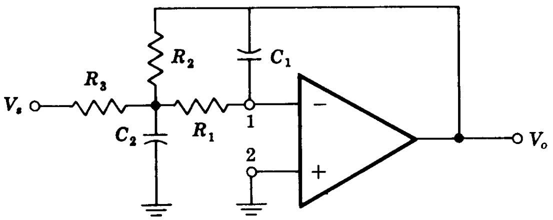
$$P_3(+s)P_3(-s)\Big|_{s=j\omega} = 1 + \omega^6$$

16-25 Show that the voltage gain $A_V(s) = V_o/V_s$ in Fig. 16-18*a* is given by Eq. (16-24). HINT: Use feedback principles.

16-26 Design an active sixth-order Butterworth low-pass filter with a cutoff frequency (or upper 3-dB frequency) of 1 kHz.

16-27 The circuit shown uses an ideal OP AMP.
(*a*) Find the voltage gain $A_V = V_o/V_s$, the damping factor k, and the cutoff frequency ω_o.
(*b*) Using this circuit, design a second-order Butterworth low-pass filter with $f_o = 1$ kHz and low-frequency voltage gain equal to -1.

Prob. 16-27



16-28 Define the z parameters of a two-port network by the following relations:

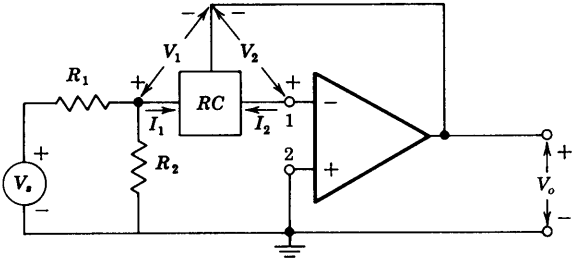
$$V_1 = z_{11}I_1 + z_{12}I_2$$
$$V_2 = z_{21}I_1 + z_{22}I_2$$

For the circuit shown prove that the voltage gain $A_V = V_o/V_s$ is given by

$$\frac{V_o}{V_s} = -\frac{R_2z_{21}}{(R_1 + R_2)(z_{11} - z_{21}) + R_1R_2}$$

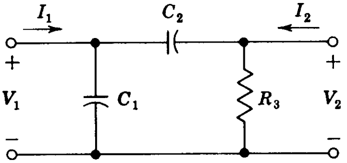
where z_{11} and z_{21} are the z parameters of the RC network.

Prob. 16-28



16-29 The network shown is the RC network of Prob. 16-28.
(*a*) Find the parameters z_{11} and z_{21} of this network.
(*b*) Find the voltage gain V_o/V_s of the amplifier in Prob. 16-28 if this RC network is used.

Prob. 16-29



16-30 Design a second-order active RC bandpass filter that has a midband voltage gain of 40 dB, center frequency of 100 Hz, and no specified bandwidth. However, the circuit must provide at least 20-dB rejection one decade from the center frequency and hold phase shift to ±10° maximum for 10 percent change from the center frequency. HINT: Use Fig. 16-22 to find Q and let $R_2 = \infty$, $R_1 = 1$ K.

16-31 Design a bandpass RC active filter with midband voltage gain of 30, center frequency of 200 Hz, and $Q = 5$. HINT: Choose $C_1 = C_2 = 0.1\ \mu$F.

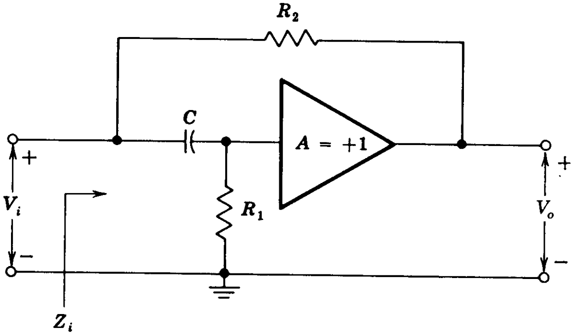
16-32 Design the resonant RLC bandpass filter of Fig. 16-21 with $f_o = 160$ Hz, 3-dB bandwidth $B = 16$ Hz, and minimum input resistance seen by the voltage source V_s of 1,000 Ω. Is this a practical circuit?

16-33 Verify Eq. (16-50) for the transfer function of the delay equalizer of Fig. 16-24*a*.

16-34 (*a*) Show that the circuit of the accompanying figure can simulate a grounded inductor if $R_1 > R_2$. In other words, show that the reactive part of the input impedance of this circuit is positive if $R_1 > R_2$.

(*b*) Find the frequency range in which the $Q = \omega L/R$ of the inductor is greater than unity.

Assume that the unity gain amplifier has infinite input resistance and zero output resistance.



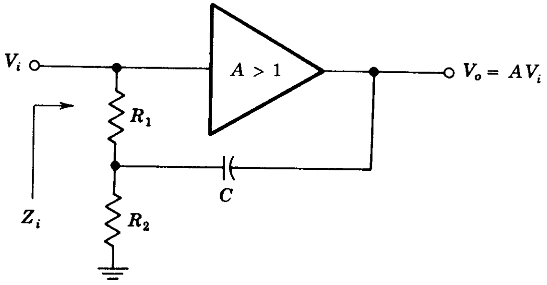
Prob. 16-34

16-35 (*a*) Show that the circuit of the given figure can simulate a grounded inductor if $A > 1$. In other words, show that the reactive part of Z_i is positive.

(*b*) Show that the real part of Z_i becomes zero ($Q = \infty$) at the frequency

$$\omega = \frac{1}{R_2 C}\sqrt{\frac{R_1 + R_2}{R_1(A - 1)}}$$

Assume that the input resistance of the amplifier of gain A is infinite.



Prob. 16-35

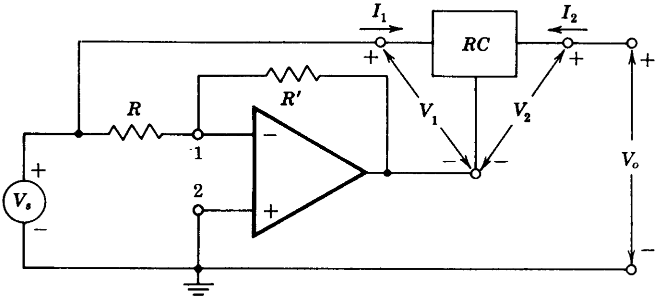
16-36 Using Fig. 16-27*b*, derive Eqs. (16-54) and (16-55).

16-37 Using Fig. 16-27*b*, derive Eqs. (16-57) and (16-58).

16-38 The figure shows a circuit using an ideal OP AMP and an *RC* two-port network. The *RC* two-port is defined in terms of its *y* parameters (Sec. 16-9). Show that the voltage gain $A_V = V_o/V_s$ is given by

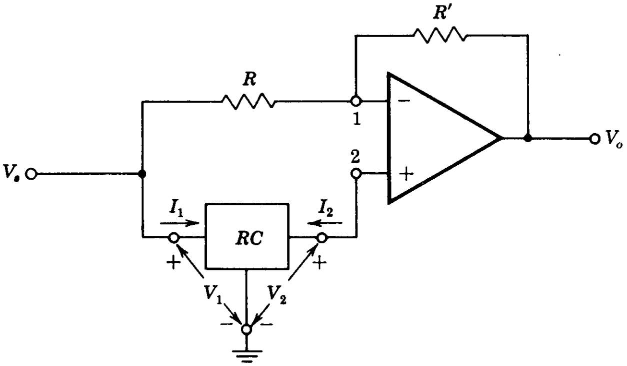
$$A_V = \frac{V_o}{V_s} = -\frac{y_{21}(1+k) + ky_{22}}{y_{22}} \qquad \text{where } k = \frac{R'}{R}$$

Prob. 16-38



16-39 Repeat Prob. 16-38 for the circuit shown. Show that the expression for $A_V = V_o/V_s$ is the same as in Prob. 16-38.

Prob. 16-39

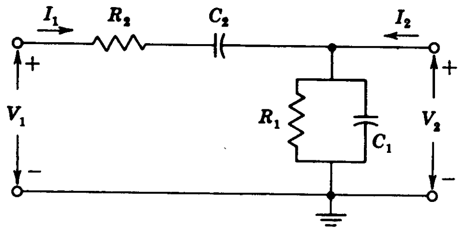


16-40 In Probs. 16-38 and 16-39 the *RC* two-port shown is used. (*a*) Find the parameters y_{21} and y_{22} of this two-port *RC* network. (*b*) Show that if

$$\frac{1}{k} = 2\left(\frac{R_2}{R_1} + \frac{C_1}{C_2}\right) + 1$$

then the two circuits are delay equalizers with transfer function

$$A_V = \frac{V_o}{V_s} = A_{V_o}\frac{(s-s_1)(s-s_2)}{(s+s_1)(s+s_2)}$$



Prob. 16-40

16-41 Using the curve of Fig. 16-31 and assuming $V_{AGC} = 3.5$ V, with an audio modulating signal of 1.5 V peak to peak, calculate the modulation factor

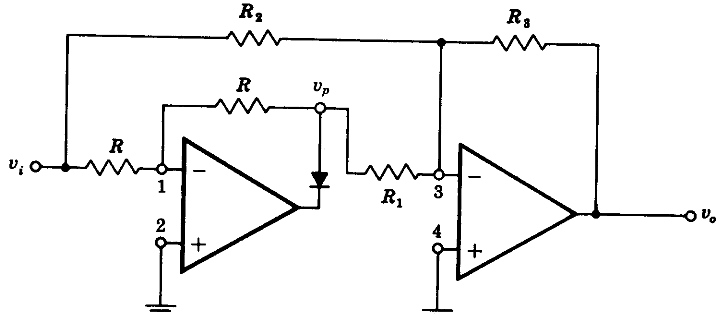
$$k = \frac{V_{o,\max} - V_{o,\min}}{V_{o,\max}}$$

16-42 Derive Eq. (16-59) for the gain of the cascode video amplifier.

16-43 (*a*) Verify that the circuit shown gives full-wave rectification provided that $R_2 = 2R_1$.

(*b*) What is the peak value of the rectified output?

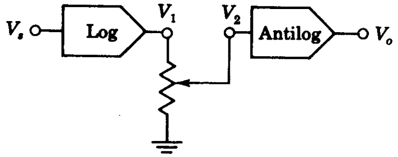
(*c*) Draw carefully the waveforms $v_i = 10 \sin \omega t$, v_p, and v_o if $R_3 = 2R_1$.



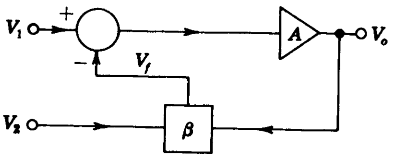
Prob. 16-43

16-44 If a waveform has a positive peak of magnitude V_1 and a negative peak of magnitude V_2, draw a circuit using two peak detectors whose output is equal to the peak-to-peak value $V_1 - V_2$.

16-45 Show that the given circuit can be used to raise the input V_s to an arbitrary power. Assume $V_1 = K_1 \ln K_2 V_s$, $V_o = K_3 \ln^{-1} K_4 V_2$, $V_2 = \alpha V_1$.



Prob. 16-45



Prob. 16-46

16-46 For the feedback circuit shown, the nonlinear feedback network β gives an output proportional to the product of the two inputs to this network, or $V_f = \beta V_2 V_o$. Prove that if $A = \infty$, then $V_o = KV_1/V_2$, where K is a constant.

16-47 (*a*) With the results of Prob. 16-46, draw the block diagram of a system used to obtain the square root of the voltage V_s.

(*b*) What should be the value of β if it is required that $V_o = \sqrt{V_s}$?

16-48 (*a*) Verify Eq. (16-78) for the pulse width of a monostable multivibrator.

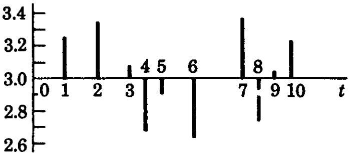
(*b*) If $V_z \gg V_1$ and $\beta = 1/2$, what is T?

16-49 Verify Eq. (16-84) for the frequency of the triangle waveform.

16-50 The Schmitt trigger of Fig. 16-47 is modified to include two clamping Zener diodes across the output as in Fig. 16-45*a*. If $V_z = 4$ V and $A_v = 5{,}000$ and if the threshold levels desired are 6 ±0.5 V, find (*a*) R_2/R_1, (*b*) the loop gain, and (*c*) V_R. (*d*) Is it possible to set the threshold voltage at a negative value? (*e*) In part (*a*) the ratio of R_2 to R_1 is obtained. What physical conditions determine the choice of the individual resistances?

16-51 The input v_i to a Schmitt trigger is the set of pulses shown. Plot v_o versus time. Assume $V_1 = 3.2$ V, $V_2 = 2.8$ V, and $v_o = +5$ V at $t = 0$.

Prob. 16-51



16-52 (*a*) Calculate the logic levels at output Y of the ECL Texas Instruments gate shown. Assume that $V_{BE,\text{active}} = 0.7$ V. To find the drop across an emitter follower when it behaves as a diode assume a piecewise-linear diode model with $V_\gamma = 0.6$ and $R_f = 20\ \Omega$.

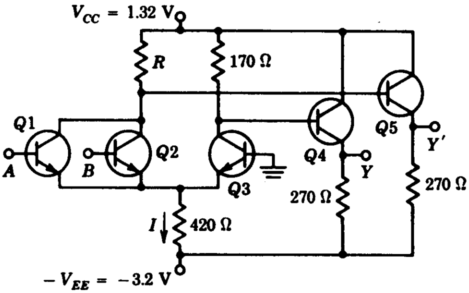
(*b*) Find the noise margin when the output Y is at $V(0)$ and also at $V(1)$.

(*c*) Verify that none of the transistors goes into saturation.

(*d*) Calculate R so that $Y' = \bar{Y}$.

(*e*) Find the average power taken from the power source.

Prob. 16-52



16-53 Verify that, if the outputs of two (or more) ECL gates are tied together as in Fig. 16-51, the OR function is satisfied.

16-54 (*a*) For the system in Fig. 16-51 obtain an expression for Y which contains three terms.

(*b*) If in Fig. 16-51 $\bar{Y}_1$ and $\bar{Y}_2$ are tied together, verify that the output is $Y = \bar{A}\bar{B} + \bar{C}\bar{D}$.

(*c*) If in Fig. 16-51 Y_1 and Y_2 are tied together and if the input to the lower ECL gate is $\bar{C}$ and $\bar{D}$ (instead of C and D), what is Y?

CHAPTER 17

17-1 Indicate how to implement S_n of Eq. (17-1) with AND, OR, and NOT gates.

17-2 Verify that the sum S_n in Eq. (17-1) for a full adder can be put in the form

$$S_n = A_n \oplus B_n \oplus C_{n-1}$$

17-3 (*a*) For convenience, let $A_n = A$, $B_n = B$, $C_{n-1} = C$, and $C_n = C^1$. Using Eq. (17-4) for C^1, verify Eq. (17-5) with the aid of the Boolean identities in Table 6-4; in other words, prove that

$$\bar{C}^1 = \bar{B}\bar{C} + \bar{C}\bar{A} + \bar{A}\bar{B}$$

(*b*) Evaluate $D \equiv (A + B + C)\bar{C}^1$ and prove that S_n in Eq. (17-1) is given by

$$S_n = D + ABC$$

17-4 (*a*) Verify that an EXCLUSIVE-OR gate is a true/complement unit.
(*b*) One input is A, the other (control) input is C, and the output is Y. Is $Y = A$ for $C = 1$ or $C = 0$?

17-5 For the system shown in Fig. 17-11*a*, verify the truth table in Fig. 17-11*b*.

17-6 (*a*) Make a truth table for a binary half subtractor A minus B (corresponding to the half adder of Fig. 17-3). Instead of a carry C, introduce a *borrow* P.
(*b*) Verify that the digit D is satisfied by an EXCLUSIVE-OR gate and that P follows the logic "B but not A."

17-7 Consider an 8-bit comparator. Justify the connections $C' = C_L$, $D' = D_L$, and $E' = E_L$ for the chip handling the more significant bits. HINT: Add 4 to each subscript in Fig. 17-14. Extend Eq. (17-12) for E and Eq. (17-13) for C to take all 8 bits into account.

17-8 (*a*) By means of a truth table verify the Boolean identity

$$Y = (A \oplus B) \oplus C = A \oplus (B \oplus C)$$

(*b*) Verify that $Y = 1(0)$ if an odd (even) number of variables equals 1. This result is *not* limited to three inputs, but is true for any number of inputs. It is used in Sec. 17-3 to construct a parity checker.

17-9 Construct the truth table for the EXCLUSIVE-OR tree of Fig. 17-15 for all possible inputs A, B, C, and D. Include $A \oplus B$ and $C \oplus D$ as well as the output Z. Verify that $Z = 1(0)$ for odd (even) parity.

17-10 (*a*) Draw the logic circuit diagram for an 8-bit parity check/generator system.
(*b*) Verify that the output is 0(1) for odd (even) parity.

17-11 (*a*) Verify that if $P' = 1$ in Fig. 17-15, this system is an even-parity check. In other words, demonstrate that with $P' = 1$, the output is $P = 0(1)$ for even (odd) parity of the inputs A, B, C, and D.
(*b*) Also verify that P generates the correct even-parity bit.

17-12 (*a*) Indicate an 8-bit parity checker as a block having 8 input bits (collectively designated A_1), an output P_1, and an input control P_1'. Consider a